Characterisation of the rif and stevor Multigene Families in Plasmodium falciparum Isolates Sampled From Natural Infections

Thesis

CHARACTERISATION OF THE *RIF* AND *STEVOR* MULTIGENE FAMILIES IN PLASMODIUM FALCIPARUM ISOLATES SAMPLED FROM NATURAL INFECTIONS.

A thesis submitted for the degree of Doctor of Philosophy

Open University

Affiliated Research Centre

KEMRI-Wellcome Trust Research Programme, Kilifi, Kenya

Collaborating Establishments

Wellcome Trust Sanger Institute, Cambridge, UK

Priscilla Wairimu Ngotho

August 2015

DATE of SUBMISSION : 13 AUGUST 2015

DATE of AWARD : 25 JANUARY 2016

ABSTRACT

Parasite-derived variant surface antigens (VSA) expressed on the infected erythrocyte surface are important targets of naturally acquired protective immune responses against malaria. The three major VSA-encoding multigene families of *Plasmodium falciparum*, *var*, *rif* and *stevor*, exhibit high inter- and intra-strain genomic variability. The VSA component encoded by the *var* gene family, PfEMP1, comprises the major cytoadhesive ligand and the major antigenic target of protective antibodies. However, other VSA families, including those encoded by the *rif* and *stevor* multigene families, may also play important roles in malaria pathogenesis and immunity. However, the biological relevance of non-PfEMP1 VSA remains largely unknown.

This thesis represents the first extensive analysis of sequence diversity and expression patterns of *rif* and *stevor* variant gene families in African field isolates of *P. falciparum*. The work details the characterization of *rif* and *stevor* gene repertoires of *P. falciparum* parasites from diverse geographical locations and identification of conserved genes in both *P. falciparum* and the related chimpanzee malaria parasite *P. reichenowi.*

Both capillary sequencing and clone-free, 454 deep sequencing methods have been used to study changes in variant gene expression during asexual development of wild and culture-adapted isolates. Isolate- and stage-specific transcription patterns of *rif* and *stevor* genes were observed, and the major sets of transcripts in multiple isolates, including parasites that have been selected for different cytoadhesive phenotypes (rosetting and adhesion to human brain endothelial cell lines) identified.

The role of RIFIN antigens in the natural acquisition of antibodies to malaria was tested using a strain transcendent variant PF3D7_1253700 (PFL2585c). The hypervariable region of this RIFIN, expressed as a recombinant protein was used to purify naturally acquired human antibodies from sera from malaria-immune African adults. These antibodies recognized native RIFIN antigen on the surface of intact trophozoites showing that RIFINs for additional targets of naturally acquired antibodies that recognize the surface of parasite-infected red blood cells.

ACKNOWLEDGEMENTS

I would like to express my greatest appreciation to my supervisor Vandana Thathy for her dedication and guidance throughout my PhD, for giving me the opportunity to work with her in this exciting project, and critical review of this thesis. My sincere gratitude also goes to my director of studies Kevin Marsh for the support, mentorship, and encouragement, and for critically reviewing my thesis. I would like to thank Chris Newbold for insightful input in my PhD.

Special thanks to Vandana Thathy's group members; Elizabeth Kibwana for keeping our group organized and infusing fun into lab work. Kadryn, Dorothy, Domtila and Vittoria Offedu for all their support.

I am particularly grateful to our collaborators, Pete Bull, Margaret Mackinnon, Etienne de Villiers, Matthew Berriman, Thomas Otto, Martin Hunt, Caroline Buckee and Daniel Larremore.

Sincere gratitude to colleagues at the KEMRI-Wellcome Trust Research Programme; Cheryl, Martin, Abdi, Opiyo, Caroline, George, Jennifer, Moses, Barnes, Ester, Mary, Joyce, Joan and all members of the immunology department for assistance in and out of the lab. Special thanks goes to Greg Fegan for help with statistical analysis. Thanks to the training department, Sam Kinyanjui and Liz Murabu, and to James Nokes my PhD third party monitor.

Finally my greatest gratitude goes to mom and dad for your unrelenting love and support, my brothers, sister and nephew for being the greatest, my mentor, aunt Charity for everything.

To my friends and support system Jackie, Sly, Faiza and Victor, I can never thank you enough.

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

2TM	Two-transmembrane
BAM	Binary sequence Alignment/Map format
BCS	Blantyre Coma Score
BLAST	Basic Local Alignment Search Tool
BSA	Bovine Serum Albumin
CDS	Coding Sequences
CIDR	Cystein-rich Interdomain Region
CM	Cerebral Malaria
CNV	Copy Number Variation
DAPI	4',6-diamidino-2-phenylindole
DBL	Duffy-binding Like
DIG	Digoxygenin
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
FITC	Fluorescein Isothiocyanate
GTR	Generalized Time Reversible
HBEC	Human Brain Endothelial Cells
HVR	Hypervariable Region
IC	Impaired Conciousness
IDC	Intraerythrocytic Developmental Cycle
IE	Infected Erythrocyte

IFA	Immunofluorescent assay
IgG	Immunoglobulin G
KEMRI	Kenya Medical Research Institute
KWTRP	KEMRI Wellcome Trust Research Programme
MC	Maurer's Clefts
MID	Multiplex Identifiers
MM	Mild Malaria
MOPS	3-morpholinopropane-1-sulfonic acid
MPC	Magnetic Particle Concentrator
mRNA	messenger Ribonucleic acid
MSP	Merozoite surface Protein
MUSCLE	Multiple Sequence Comparison by Log-Expectation
NaAC	Sodium acetate
NCBI	National Center for Biotechnology Information
NGS	Next Generation Sequencing
NS	Non-severe malaria
PBMC	Peripheral Blood Mononuclear Cell
PBS	Phosphate Buffered saline
PCR	Polymerase Chain Reaction
PEXEL	Plasmodium Export Element
PfEMP1	*P. falciparum* Erythrocyte Membrane Protein 1
PTEX	Plasmodium Translocon of Exported Proteins
PV	Parasitophorous Vacuole

PVM	Parasitophorous Vacuole Membrane
QRT-PCR	Quantitative Real Time Polymerase Chain Reaction
R-	Non-rosetting
R+	Rosetting
RT-PCR	Reverse Transcriptase PCR
RBC	Red Blood Cell
RD	Severe Respiratory Distress
RF	Rosetting frequency
rif	Repetitive Interspersed repeat family
RNA	Ribonucleic acid
RPKM	Reads Per Kilo basepair per Million
rRNA	ribosomal Ribonucleic acid
SAM	Sequence Alignment/Map format
SDI	Simpson Diversity Index
SICA	Schizont-Infected Cell Agglutination
SMA	Severe Malarial Anaemia
SPRI	Solid Phase Reversible Immobilization
stevor	subtelomeric variant open reading frame
TAE	Tris Acetate EDTA
TE	Tris EDTA
TM	Transmembrane
TVN	Tubulovesicular Network
VSA	Variant Surface Antigens

WGS	Whole Genome Sequence
WHO	World Health Organisation
WTSI	Wellcome Trust Sanger Institute

CONTENTS

CHAPTER 1

1 LITERATURE REVIEW

1.1 Malaria – an overview

Although malaria is a preventable and curable disease, it still remains a major public health problem causing significant morbidity and mortality. In the year 2013, an estimated 198 million malaria cases were reported, of which 584,000 resulted in death. Of these deaths, 90% occurred in sub-Saharan Africa (SSA) (WHO report, 2014). Malaria is a vector-borne disease caused by Apicomplexan parasites of the genus *Plasmodium,* which have been found to infect mammals, reptiles and birds, and are transmitted by a variety of insect vectors. *Anopheles* mosquito vectors transmit *Plasmodium* species that infect humans and other mammals, whereas *culicine* mosquitoes (*Aedes* and *Culex* sp.) are primarily responsible for the transmission of avian plasmodia. All *Plasmodium* species show two distinct types of division, asexual division and sexual division. There are two forms of asexual division, schizogony within the vertebrate host and sporogony in the insect vector. Sexual division occurs solely in the definitive insect host. Four species of *Plasmodium* naturally cause malaria in humans, *P. falciparum, P. vivax, P. malariae* and *P. ovale.* Recently a fifth species *P. knowlesi* has been shown to be a cause of malaria in humans (Singh et al., 2004). The infection is zoonotic in nature and it is not known whether human-to-human transmission occurs. The most lethal form of malaria is caused by *P. falciparum,* which is the basis of this thesis.

1.2 *P. falciparum* malaria

P. falciparum malaria remains a major public health problem in tropical countries especially in SSA where the burden of disease is borne by young children and pregnant women. Currently, there are no licensed vaccines. The need for effective malaria vaccines is justified by the high cost of early childhood mortality as well as the epidemiological transition of some areas of high transmission to low transmission (Okiro et al., 2007, Ceesay et al., 2008, Kleinschmidt et al., 2009, O'Meara et al., 2008) leading to large populations of less immune individuals. In addition widespread resistance of parasites to drugs and vectors to insecticides complicates alternative malaria control efforts.

1.2.1 Life cycle of *Plasmodium falciparum*

P. falciparum has a complex life cycle involving two hosts (Figure 1.1). The parasite life cycle begins when a female *Anopheles* mosquito takes a blood meal from an infected individual containing intra-erythrocytic sexual forms termed gametocytes (Figure 1.1. 8, 9, and 10). Gametocytes, the precursors of male and female gametes are crescent-shaped forms containing a single nucleus that persist within erythrocytes in the periphery as mature forms arrested in G_0. In the mosquito gut, changes in the environment including a drop in temperature, an increase in carbon dioxide and other mosquito factors trigger differentiation of gametocytes to gametes. The gametocytes escape the host cell by lysing the infected erythrocyte. The male gametocytes simultaneously go through three mitotic divisions resulting in 8 male gametes, a process known as exflagellation (Figure 1.1. 11). The male and female gametes fuse to form a diploid zygote. Within 20-24 hours, the

zygote elongates and differentiates to a highly motile, invasive form known as an ookinete (Figure 1.1. 12). The ookinete escapes from the blood meal, and invades the midgut epithelial wall, lodging underneath the basal lamina, where it differentiates into a spherical oocyst (Figure 1.1.13). Within 8 to 15 days the oocyst matures as multiple mitotic nuclear divisions occur, producing thousands of haploid sporozoites. When the process of sporozoite budding (sporogony) is complete, the oocyst ruptures releasing the sporozoites (Figure 1.1. 1 and 14) into the hemolymph, a fluid that bathes all the organs in the mosquito's body cavity. Sporozoites are passively carried to the salivary glands, where they bind to and penetrate the basal lamina, invade the salivary gland epithelial cells, exit into secretory cavities and only a minor proportion eventually enter the salivary ducts to await transmission (Pimenta et al., 1994, Frischknecht et al., 2004). During feeding about 1-100 sporozoites are injected into the subcutaneous tissue of an individual together with the mosquito's saliva (Medica and Sinnis, 2005, Jin et al., 2007). Some sporozoites remain in the skin or invade lymphatic vessels and are subsequently degraded (Amino et al., 2006). Some sporozoites enter blood vessels and flow into the liver where they traverse through several hepatocytes and Kupffer cells (Vanderberg et al., 1990, Mota et al., 2001, Mota et al., 2002, Mota and Rodriguez, 2004, Menard, 2001) before invading a hepatocyte and differentiating into a clinically silent liver stage schizont (Figure 1.1. 2 and 3). After 5 to 7 days, merozoite filled vesicles termed merosomes (Figure 1.1. 4) bud off from the mature schizont into the blood stream (Sturm et al., 2006, Graewe et al., 2011). Thousands of merozoites are liberated into the blood stream by the breakdown of the host cell plasma membrane and parasitophorous vacuole membrane (PVM) that enclose the merosomes (Graewe et al., 2011). The newly released merozoites

invade red blood cells (RBCs) (Figure 1.1. 5) and establish the clinically relevant, pathogenic, intraerythrocytic developmental cycle (IDC), where multiple cycles of erythrocytic schizogony occur. Some merozoites that invade erythrocytes do not go through asexual erythrocytic schizogony but develop into the sexual blood stages, male or female gametocytes (Dyer and Day, 2000, Smith et al., 2002, Drakeley et al., 2006, Bousema and Drakeley, 2011). These transmissible forms are taken up by the mosquito to complete the parasite life cycle.

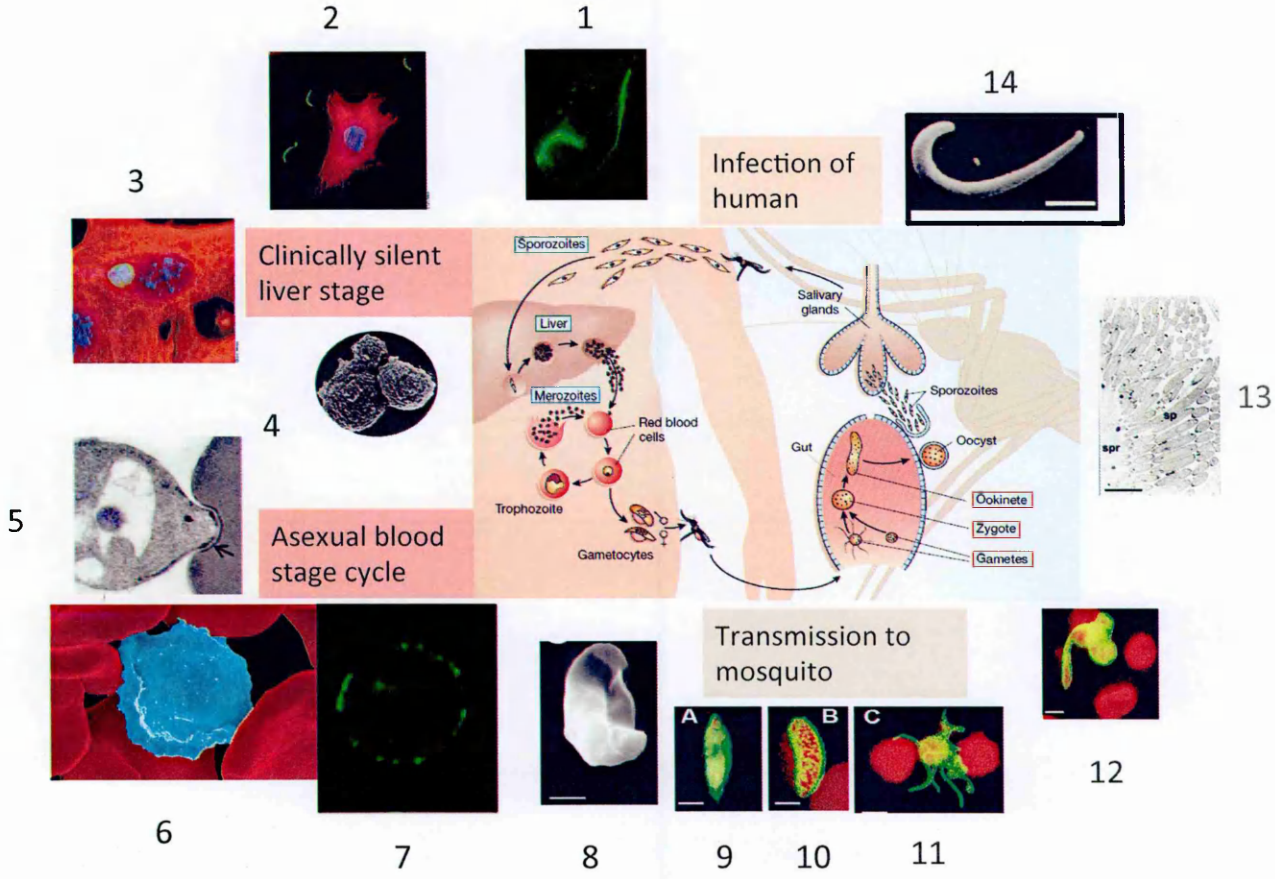

1
2
3
4
5
6
7
8
9
10
11
12
13
14
Infection of human
Clinically silent liver stage
Asexual blood stage cycle
Transmission to mosquito
Sporozoites
Liver
Merozoites
Red blood cells
Trophozoite
Gametocytes
Salivary glands
Sporozoites
Gut
Oocyst
Ookinete
Zygote
Gametes
sp
spr
A
B
C

Figure 1.1. *Life cycle of P. falciparum. Image from adapted from Menard R. (Menard, 2005)*

1: P. falciparum sporozoite, air-dried, and labeled with hyperimmune Kenyan adult sera, followed by secondary fluorescent antibody (AlexaFluor488 goat anti-human IgG). The sporozoite's surface coat and the trail produced as it glides are shown in green. Photo by Offeddu, V. and Thathy V, unpublished.

2: Sporozoites migrating to hepatocytes to initiate liver stage infection. Parasites in green, tubulin in red and DNA in blue; Photo by Charo Armas (Preiser, 2002)

3: A multinucleated liver stage schizont developing within a vacuole in the hepatocytes; Photo by Charo Armas (Preiser, 2002).

4: Scanning electron micrograph of a budding merosome (Sturm et al., 2006).

5: Transmission electron micrograph of merozoite invading an erythrocyte (Srinivasan et al., 2011).

6: Colourized electron micrograph of an infected erythrocyte (blue), which has altered morphology and permeability pathways as compared to the uninfected erythrocytes (red): Photo by A. Mora (Sanchez, 2011).

7: Live immunofluorescence assay showing an IE surface stained with anti-stevor serum (Niang et al., 2009).

8: Scanning electron micrograph of a mature gametocyte (Pradel, 2007).

9: Stage IV gametocyte showing surface-associated Pfs230 expression (shown in green) (Pradel, 2007).

10: Stage V gametocyte surface showing gametocyte-specific multi-domain adhesion protein PfCCp1 (in green, erythrocyte shown in red) (Pradel, 2007).

11: Exflagellating microgametocyte with 8 microgametes expressing alpha-tubulin II (alpha-tubulin II labeling shown in green, erythrocytes shown in red) (Pradel, 2007).

12: A cluster of 3 zygotes, one of which is transforming into an ookinete, expressing Pfs25 on the surface (Pfs25 labeling shown in green, erythrocytes shown in red) (Pradel, 2007)

13: Electron micrograph of an oocyst packed with sporozoites (sp) budding off the sporoblast (spr) (Thathy et al., 2002).

14: Scanning electron micrograph of midgut sporozoites isolated from infected mosquitoes on day 14 post infection. Bar: 2 um. (Thathy et al., 2002).

1.2.2 Intraerythrocytic developmental cycle (IDC)

Invasion of host erythrocytes as well as other cells is possible mainly due to an apical complex (anterior pole) found in infective forms, the "zoites" of members of the Apicomplexan phylum (Baum et al., 2008, Sinnis and Sim, 1997). In the case of *Plasmodium*, the merozoite invades erythrocytes to initiate the IDC, a clinically important phase of the parasite's life cycle where the symptoms associated with malaria morbidity and mortality are manifested (Figure 1.1. 5, 6 and 7). Erythrocyte invasion is a complex receptor-mediated process involving an array of factors, including proteins from the apical secretory organelles, micronemes, rhoptries, and dense granules. Invasion involves binding of a merozoite to the erythrocyte, reorientation, formation of a tight junction (Figure 1.1. Image 5), and entry into the erythrocyte (Riglar et al., 2011). After invasion, the parasite goes through a period of intracellular growth, followed by asexual replication by schizogony, erythrocyte rupture and release of multiple daughter merozoites 24-72 hours later, depending on the species. *P. falciparum* undergoes a 48-hour growth cycle from the initial ring form, to trophozoite and finally the schizont stage, which ruptures to release up to 32 daughter merozoites (Figure 1.1). The newly released merozoites invade other erythrocytes perpetuating the growth and asexual division cycle until either the host dies or the parasite is controlled by the host's immunity or drug therapy.

Inside the erythrocyte the parasite exists inside a membrane bound vacuole, the parasitophorous vacuole (PV). The parasites ingest the erythrocyte cytosol and digests hemoglobin in the food vacuole to obtain amino acids. Free heme is detoxified by converting into the malaria pigment hemozoin. Erythrocytes are metabolically inert host

cells. Thus, after invasion, the parasite modifies the infected erythrocyte (IE) by formation of new permeability pathways (Ginsburg and Stein, 2004, Ginsburg and Stein, 1987, Ginsburg et al., 1983) to allow influx of low molecular weight metabolites as well as set up elaborate tubular structures, the tubulovesicular network (TVN) and Maurer's clefts (MC) (Tilley et al., 2008) for protein export to the IE surface to interface with the host (Spielmann et al., 2006). Parasites dramatically remodel the IE by exporting hundreds of proteins, the exportome to enable survival and proliferation in the bloodstream (Maier et al., 2009). The exported proteins alter cellular properties of the host cell making them less deformable, and enables expression of parasite antigens on the surface of IE (Maier et al., 2009). The surface proteins alter the antigenic properties of the IE and undergo clonal antigenic variation to evade host immunity. In addition surface antigens mediate adhesion of the late-stage-IE to a variety of cell receptors leading to sequestration of the IE in the microvasculature of different organs. It is likely that tissue specific sequestration of IE plays a role in the pathogenesis of different clinical phenotypes, such as cerebral malaria associated with parasitized red cells binding to brain endothelial cells (Turner et al., 2013, Moxon et al., 2013) and pregnancy malaria, associated with parasites sequestering in the placenta (Duffy and Fried, 2003b, Beeson et al., 2001). Late-stage IE can also adhere to uninfected erythrocytes forming "rosettes". Rosetting is believed to contribute to sequestration of IE through mechanical obstruction of the capillaries (Mercereau-Puijalon et al., 2008)

1.3 Malaria pathogenesis

P. falciparum infection in a non-immune individual is most likely to result in clinical disease. During clinical disease, symptoms including fever spasms, muscle aches and headaches, are manifested during the asexual erythrocytic cycle. In *P. falciparum* the spasms of fever occur every 48 hours, which coincides with the rupture of schizont IE and release of merozoites, parasite antigens, and pigment into the bloodstream. Clinical outcomes range from asymptomatic infections to severe malaria and death depending on age at infection and transmission levels (which modify host immune status). In high transmission areas incidence of clinical disease is highest in the first year of life, but as children are constantly challenged they acquire immunity to clinical malaria within the first four years of life, and thus experience fewer cases and less severe disease. In areas with low transmission the peak in disease incidence occurs at a later age (Snow et al., 1997).

Severe malaria in African children presents as a single syndrome or a combination of three overlapping syndromes, impaired consciousness, respiratory distress and severe anemia (Marsh et al., 1995). Severe malaria with respiratory distress poses the highest risk of death in African children followed by impaired consciousness. Impaired consciousness presents as deep coma and inability to make verbal or motor responses. The depth of coma is assessed based on Blantyre coma score (BCS) with a score of 4 or less classified as impaired consciousness (Molyneux et al., 1989, Newton et al., 1997, Marsh et al., 1995). A BCS of less than 3 is classified as cerebral malaria (Newton et al., 1997, Molyneux et al., 1989). Respiratory distress is characterized by recession or abnormally deep breathing with metabolic acidosis (Marsh et al., 1995). Severe malarial

anemia is defined as venous hemoglobin concentration of 5g/dl or less on admission (Marsh et al., 1995).

1.4 Immunity to malaria

Attempts to develop a malaria vaccine are based on the fact that people can acquire some degree of immunity to malaria through natural exposure or experimental vaccination. It is therefore important to understand the mechanisms of acquisition and molecular targets of natural immunity to malaria to help guide rational vaccine design. Some studies have suggested that people living in high malaria transmission areas acquire immunity to severe malaria after relatively few clinical episodes (Gupta et al., 1999). This immunity is usually acquired by the age of five years and is evidenced by decline in prevalence of clinical episodes in the immune individual. Immunity to mild febrile malaria is acquired more slowly and immunity to infection is hardly ever achieved (reviewed in (Marsh, 1992).

Naturally acquired immunity to malaria has stage, strain and variant specific elements (Bull and Marsh, 2002, Langhorne et al., 2008, Doolan et al., 2009) but eventually the individual may develop more cross protective responses (Marsh and Howard, 1986, Langreth and Reese, 1979, Hommel et al., 1983). Antibody responses to blood stage parasites play an important role in the development of protective immunity to malaria (Cohen et al., 1961, Marsh et al., 1989, Bull et al., 1998, Duffy and Fried, 2003a, Staalsoe et al., 2004). Studies on passive transfer of gamma globulin from malaria immune adults to infected children showed that *P. falciparum*-induced antibodies rapidly reduced

parasitemia (Cohen et al., 1961). In addition maternal transfer of antibodies has been shown to protect infants against malaria (Edozien, 1962).

Antibodies can be targeted to the free merozoites or IE. Immune responses to the IE are targeted to parasite derived variant surface antigens (VSA) expressed on the IE (Marsh et al., 1989, Newbold et al., 1992). A number of studies have suggested that antibodies against these VSA confer protection against malaria (Marsh et al., 1989, Fried and Duffy, 1998, Bull et al., 1998, Giha et al., 2000, Dodoo et al., 2001).

1.5 Variant Surface Antigens (VSA).

During the asexual growth cycle malaria parasites reside inside erythrocytes, cells that lack surface major histocompatibility complex (MHC) molecules and hence have no capacity for antigen presentation. Thus the parasites would stay hidden from the host immune responses. The hypothesis that the parasites were "invisible" to host immune system was challenged with the advent of the schizont-infected cell agglutination assay (SICA) (Eaton, 1938). Using the SICA test in a monkey malaria model, Eaton demonstrated that serum from an infected monkey agglutinated *P. knowlesi* IE (Eaton, 1938). This agglutination was species specific and was observed when erythrocytes infected with mature stage parasites were incubated with serum. Uninfected erythrocytes as well as early-stage-IE were not agglutinated. Sera from monkeys that had not been infected with malaria or infected with a different species, or were acutely ill with malaria did not agglutinate IE. This suggested that species-specific antigens were present on the surface of mature stage IE and these were targets of host immune responses. Later studies by Brown *et al.*, further explored antigenic variation in *P. knowlesi*. By infecting

monkeys with *P. knowlesi* and subjecting them to sub-curative chemotherapy, the investigators were able to induce chronic infections with each successive peak in parasitemia being lower than the previous. Each peak in parasitemia represented distinct antigenic variants expressed on IE that induced a protective immune response (Brown and Brown, 1965, Brown et al., 1968). Variants from each recrudescence were infective to naïve monkeys and induced antibodies that were able to agglutinate homologous schizont-infected erythrocytes. Sera from monkeys infected with heterologous parasites or serum collected from the same monkey before the infection did not induce agglutination. These antibodies conferred protection against homologous antigenic variants but not against heterologous variants (Brown and Brown, 1965, Brown et al., 1968), suggesting that variant antigens on mature stage IE (SICA antigens) are targets of protective immune responses. Because these studies were looking at parasite populations it was not clear whether an individual parasite was able to switch to expressing a new variant antigen within an infection. The switch could alternatively result from immune selection of the dominant variant within the population. Definitive evidence of the former came from an experiment using cloned parasites (Barnwell et al., 1983). Barnwell *et al.,* showed that multiple passages of *P. knowlesi* SICA[+] parasites through splenectomized monkeys lead to a change in phenotype to SICA[-] variants that were not agglutinatable by immune serum (Barnwell et al., 1982). They later cloned both SICA[+] and SICA[-] by micromanipulation. The cloned SICA[+] parasites switched to SICA[-] after multiple passages in splenectomized monkeys and the SICA[-] to SICA[+] when passaged through intact monkeys (Barnwell et al., 1983). This conclusively showed an individual parasite had the capacity to switch between different antigenic variants.

Following these studies the major *P. knowlesi* surface antigen was identified by immunoprecipitating radiolabelled antigens from schizont-IE from cloned parasites of distinct SICA phenotypes using immune sera. These antigens were characterized as high molecular weight parasite proteins that were expressed on schizont-IE in a variant specific manner (Howard et al., 1983). Similarly experiments using in vitro cultures of *P. falciparum* identified a high molecular weight protein as the major VSA on the surface of trophozoite-IE (Leech et al., 1984, McBride et al., 1985) and later termed *P. falciparum* erythrocyte membrane protein 1 (PfEMP1) (Baruch et al., 1995).

Clonal antigenic variation in VSAs in *P. falciparum* was demonstrated *in vitro* using cloned parasites from the isolate IT (Biggs et al., 1991, Roberts et al., 1992).

These studies suggest that targets of naturally acquired antibodies to malaria are highly diverse and undergo switching between variants, suggesting that immunity to malaria may involve acquisition of a wide repertoire of antibodies against multiple polymorphic antigens. This might be the reason why immunity to mild malaria is acquired over many years and immunity to infection hardly ever achieved (Marsh, 1992). On the other hand immunity to severe malaria is acquired relatively quickly after a few symptomatic episodes (Gupta et al., 1999). In addition, epidemiological studies have shown that people living in malaria endemic areas develop immunity to severe malaria early in life with older children and adults being resistant to severe malaria (Marsh and Kinyanjui, 2006). This suggests that there exist immune targets with restricted heterogeneity.

The importance of VSA as targets of naturally acquired immunity to malaria was suggested in a longitudinal study of children living in a malaria endemic area in The Gambia, which showed that the level preexisting antibodies to the IE surface was the best

predictor of protection from subsequent clinical episodes (Marsh et al., 1989). Later studies showed that children develop anti-VSA antibodies to the variants that caused an acute episode (Marsh and Howard, 1986, Newbold et al., 1992, Bull et al., 1999), and acquisition of antibodies to a specific antigenic variant protected the children from clinical episodes caused by infection with that variant type (Bull et al., 1998, Giha et al., 2000).

Three major multigene families present in the *P. falciparum* genome (Gardner, 2002) encode VSA: *var* genes, which encode *P. falciparum* erythrocyte membrane protein 1 (PfEMP1), the repetitive interspersed family (*rif*) encoding RIFIN proteins, and the subtelomeric variant open reading frame (*stevor*) family, encoding STEVOR proteins (Table 1.1).

Table 1.1. *Number of genes in the multigene families encoding VSA in whole genome sequenced laboratory isolates.*

		3D7	IT	HB3	Dd2
var	No. of intact genes	63	49	38	39
	No. of pseudogenes	40	7	12	11
	Total	**103**	**56**	**50**	**50**
rif	No. of intact genes	157	121	161	99
	No. of pseudogenes	27	19	27	34
	Total	**184**	**140**	**188**	**132**
stevor	No. of intact genes	31	34	32	24
	No. of pseudogenes	9	5	11	7
	Total	**40**	**39**	**43**	**31**
pfmc-2tm	No. of intact genes	10	5	13	2
	No. of pseudogenes	3	3	1	1
	Total	**13**	**8**	**14**	**3**

3D7 v3 (completely sequenced repertoire) and IT v3 sequences obtained from GeneDB (Gardner et al., 2002).

HB3 sequences obtained from HB3 genome sequenced by the Sanger Wellcome Trust Institute (unpublished).

Dd2 sequences obtained from the Broad Institute (http://www.broad.mit.edu).

1.5.1 *var* encoded PfEMP1

The major VSA expressed on IE is the extensively characterized PfEMP1 family, encoded by the polymorphic *var* multigene family made of about 60 members (Su et al., 1995, Baruch et al., 1995, Smith et al., 1995, Gardner et al., 2002). These genes are expressed in a mutually exclusive manner (Chen et al., 1998, Scherf et al., 1998) and undergo transcription switching during intraerythrocytic growth, thus orchestrating clonal antigenic variation of the IE surface (Roberts et al., 1992, Horrocks et al., 2004). All members of this gene family share a similar structure consisting of a large 5' exon encoding a short amino-terminal segment preceding a highly variable extracellular domain, and a short 3' exon encoding the conserved intracellular acidic terminal segment (Su et al., 1995, Smith et al., 1998). The extracellular segment is made up of different combinations of Duffy Binding Like (DBL) domains and Cysteine rich Inter-Domain Regions (CIDR), which are thought to mediate cytoadhesion. Despite high sequence diversity *var* genes share key sequence features enabling classification into biologically important groups. Based on the 5' flanking, or upstream (Ups), region, most *var* genes have been classified into UpsA (group A), UpsB (group B) and UpsC (group C) (Gardner et al., 2002, Lavstsen et al., 2003). Recently, a sequence classification system based on the modular domain architecture at the 5' end of *var* genes from seven different *P. falciparum* genomes has been used to classify *var* genes into different sub-types, termed domain cassettes (DC1-DC24) (Rask et al., 2010). Importantly expression of *var* genes belonging to specific sub-types has been associated with severe malaria showing the biological significance of the classifications (Bull et al., 2000, Nielsen et al., 2002, Salanti et al., 2003, Jensen et al., 2004, Warimwe et al., 2009, Moxon et al., 2013, Turner

et al., 2013, Abdi et al., 2014). More recent work has shown that DC8 and DC13 mediate binding to brain endothelial cell lines in culture, and expression of these *var* genes is associated with severe malaria (Avril et al., 2012, Claessens et al., 2012, Lavstsen et al., 2012, Turner et al., 2013). DC4 PfEMP1 proteins have been shown to bind intracellular adhesion molecule 1 (ICAM) expressed on transformed Chinese hamster ovary cells while DC5 PfEMP1 bind platelet/ endothelial cell adhesion molecule 1 (PECAM1) expressed on transformed human bone marrow endothelial cells (Berger et al., 2013). *var* gene expression is highly controlled to ensure expression of a single PfEMP1 on the IE surface (Gardner et al., 1996). The Ups region, *var* promoter and epigenetic mechanisms are involved in *var* expression regulation (Voss et al., 2006, Voss et al., 2007, Dzikowski et al., 2006a, Frank et al., 2006, Dzikowski et al., 2007, Chookajorn et al., 2007, Gannoun-Zaki et al., 2005). The three main *var* promoters alternate between a silenced and activated state. Activation of one promoter results in silencing of all the other promoters thus silencing all other *vars*. Upregulation of an episomal *var* promoter driving expression of a drug selectable marker, led to down regulation of all endogenously expressed *var* genes, showing that these promoters act as important regulators for mutually exclusive *var* gene expression (Witmer et al., 2012).

PfEMP1 proteins are believed to play a central role in malaria pathogenesis because they can mediate adherence of the IE to various receptors on host cells (Urban et al., 1999, Rowe et al., 1997, Reeder et al., 1999, Smith et al., 1995). Further evidence of a direct link between IE adhesion and pathogenesis was provided by Moxon et al, when they observed decreased EPCR expression and increased fibrin deposition at the cytoadhesion sites

between IE and brain endothelium in post-mortem specimens from children who died from cerebral malaria (Moxon et al., 2013). PfEMP1s are also the major candidate targets of protective antibody responses directed against the surface of IE (Chan et al., 2012). However it is likely that other VSA families in *P. falciparum*, such as RIFIN and STEVOR, that may be co-expressed with PfEMP1 on the IE surface also play important roles in host-parasite interactions (Voss et al., 2006, Abdel-Latif et al., 2004, Newbold et al., 1999, Cheng et al., 1998, Kyes et al., 1999, Fernandez et al., 1999, Niang et al., 2009, Sanyal et al., 2012, Tiburcio et al., 2012, Niang et al., 2014, Goel et al., 2015). PfEMP1 is only found in *P. falciparum* and the related non-human malaria parasite *P. reichenowi*. Other *Plasmodium* species lack PfEMP1, yet they evade host immune responses and can exhibit cytoadherence phenotypes like sequestration and rosetting, suggesting that these functions are likely to be mediated by other VSA (Jemmely et al., 2010).

1.5.2 Small variant surface antigens.

RIFIN (Cheng et al., 1998, Fernandez et al., 1999, Kyes et al., 1999, Petter et al., 2007) STEVOR (Blythe et al., 2004, Limpaiboon et al., 1990, Cheng et al., 1998, Lavazec et al., 2006, Niang et al., 2009), PfMC-2TM (Sam-Yellowe et al., 2004) and SURFIN (Winter et al., 2005) have been identified as potential VSA in P. *falciparum*. They are encoded by the *rif, stevor, P. falciparum* Maurer's cleft 2- transmembrane *(Pfmc-2tm)* and *surf* multigene families respectively. The *rif, stevor* and *pfmc-2tm* genes belong to the two-transmembrane (2TM) variant gene superfamily, which share a number of structural characteristics (Figure 1.2) (Cheng et al., 1998, Lavazec et al., 2006, Sam-Yellowe et al., 2004). The *rif, stevor*, and *Pfmc-2tm* genes share a similar two-exon architecture. The

short first exon encodes a signal peptide and the longer second encodes a semi-conserved amino-terminal region, separated from a highly polymorphic region by a putative transmembrane domain, with a second predicted transmembrane domain occurring before the semi-conserved, positively charged, short carboxyl-terminal region (Figure 1.2). All members have the Plasmodium export element (PEXEL)/vacuolar transport signal (VTS) motif (Hiller et al., 2004, Marti et al., 2004) near the start of exon 2, targeting them for translocation into the host cell. The hypervariable regions of each gene family exhibit extensive sequence diversity across paralogs and between parasite isolates (Kyes et al., 1999, Lavazec et al., 2006, Joannin et al., 2008). Different topologies for the 2TM proteins on the IE surface have been proposed (Templeton, 2009b, Niang et al., 2009, Przyborski et al., 2005, Kyes et al., 1999, Bultrini et al., 2009, Goel et al., 2015, Bachmann et al., 2015). Until recently, the prevailing model was that the protein products of the 2TM superfamily possess two transmembrane (TM) domains flanking the hypervariable region, which forms a surface-exposed loop (Lavazec et al., 2006, Joannin et al., 2008) (Figure 1.3). Recent data challenge the existence of the first N-terminal putative transmembrane domain (TM1), at least for some RIFIN and STEVOR family members, and instead support a topology where both the semi-conserved N-terminal region and the hypervariable region are exposed on the IE surface and anchored to the membrane only by the C-terminal transmembrane domain (TM2) (Bultrini et al., 2009, Niang et al., 2009, Goel et al., 2015, Bachmann et al., 2015). Regardless of the exact membrane topology of the 2TM proteins, the available data suggests that the hypervariable region flanking the 2TM domains is extracellular and exposed at the host-parasite interface. The profound sequence diversity confined largely to this region, in

addition to the conserved PEXEL/VTS trafficking motif as a signature of export into the IE, has led to the hypothesis that the hypervariable region defines a surface-exposed loop that is diversifying as a result of immune selection (Sam-Yellowe et al., 2004, Dzikowski et al., 2006b, Przyborski et al., 2005, Lavazec et al., 2006, Templeton, 2009b, Kyes et al., 1999).

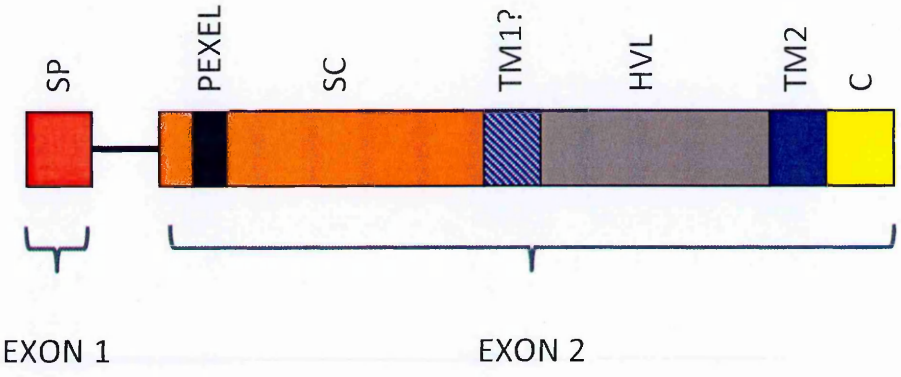


Figure 1.2. Schematic showing the architecture of *the P. falciparum 2TM variant gene superfamily. The rif, stevor, and Pfmc-2TM genes share a common two-exon structure in which a short first exon encodes a predicted signal peptide (SP) (red) and a longer exon 2 (~1kb) encodes a PEXEL/VTS export motif (black), a semi-conserved (SC) predicted globular domain (orange), two predicted transmembrane domains, a weakly predicted TM1 (shaded blue) and a strongly predicted TM2 (solid blue) which flank a hypervariable region (grey), and a short semi-conserved positively-charged carboxyl terminal region (C) (yellow). The black solid line indicates the intron. The lengths of the hypervariable regions of each protein family differ, being 170, 60 and <20 amino acids in RIFIN, STEVOR and PfMC-2TM proteins respectively*

1.5.3 The *pir* multigene family.

Unlike *var*, which is unique to *P. falciparum* and *P. reichenowi* (Otto et al., 2014b) *rif* and *stevor* belong to the larger *Plasmodium* interspersed repeat (*pir*) gene superfamily, which is found in human, monkey and rodent malarias (Janssen et al., 2004, http://www.genedb.org/genedb/malaria/, http://www.plasmodb.org/plasmo/home.jsp). *Pir* gene families are the largest in each parasite genome, and encode proteins that are expressed at the host parasite interface (Florens et al., 2002, Niang et al., 2009, Mwakalinga et al., 2012, Khattab and Meri, 2011). The *pir* multigene family has ancient evolutionary origins (Janssen et al., 2004), but the definitive function is still unknown.

There is emerging evidence of the role of PIR proteins in cytoadhesion and immune evasion (Niang et al., 2009, Bernabeu et al., 2012, Goel et al., 2015, Niang et al., 2014). In the closely related chimpanzee malaria parasite, *P. reichenowi,* the *pir* family is highly expanded in comparison to *P. falciparum.* Interestingly, the *P. reichenowi* genome contains the same three variant multi-gene families as *P. falciparum, var, rif, and stevor.* There are 568 *rif* and 66 *stevor* in the *P. reichenowi* genome, as compared to 184 *rif* and 42 *stevor* in *P. falciparum* genome *(Otto et al., 2014b).* The *pir* genes have also been identified in the human malaria parasite, *P. vivax* (*vir*) (del Portillo et al., 2001, Lopez et al., 2013), the human and simian malaria parasite *P. knowlesi* (*kir*) (Pain et al., 2008), and in rodent malaria parasites, *P. berghei (bir), P. chabaudi (cir)* and *P. yoelii* (*yir*) (Carlton et al., 2002, Otto et al., 2014a, Lawton et al., 2012). Members of the *pir* multigene family encode proteins with similar features suggesting that they might have shared functions (Janssen et al., 2004, Jemmely et al., 2010). In *P. vivax* the *vir* gene family encode VIR

proteins that are translocated to the surface of infected reticulocytes and have been associated with cytoadhesion (Carvalho et al., 2010, Bernabeu et al., 2012). A recent study on diversity of *cir* revealed similarities between *cir* and *rif* multigene families (Lawton et al., 2012). *cir* can be subdivided into two functionally divergent groups, a more diverse A subfamily and the more conserved B subfamily based on an insertion of 18 amino acids in the A subfamily. The insertion in *cirA* bears some sequence similarities to the insertion that is present in the *P. falciparum rifA* group, which is also larger and more diverse than the *rifB* sub-family (see section 1.5.3.1 and Figure 1.4 below) (Lawton et al., 2012). Importantly, and for the first time, the *cir* genes have been implicated in virulence (Spence et al., 2013). In this study, serially blood passaged (SBP) *P. chabaudi* were more virulent than mosquito transmitted (MT) parasites because of a modified immune response to the blood stage parasite. The difference in virulence between the MT and SBP parasites dependent on the presence of an intact immune system and attributed to a change in the pattern of expression of *cir* genes, with the SBP parasites expressing few *cirs*, and reaching higher parasitemias, while the MT parasites expressed many more *cir* genes and were associated with lower, more chronic parasitemias (Spence et al., 2013). This study suggests for the first time a link between *pir* gene expression, parasite multiplication rate and host immunity.

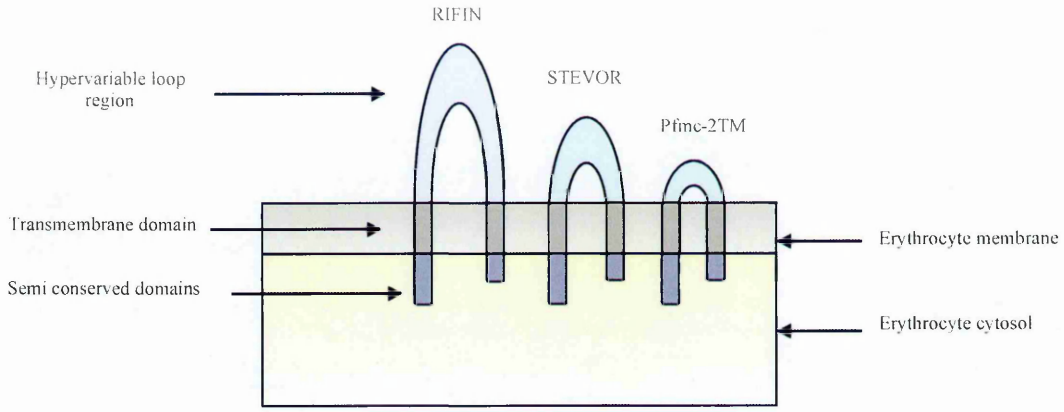


Figure 1.3. *Proposed model of the 2TM protein topology in infected erythrocytes. Figure adapted from (Scherf et al., 2008). The central hyper variable regions flanked by the two predicted transmembrane domains of RIFIN, STEVOR and PfMC-2TM proteins harbor extensive sequence polymorphism both within and across parasite isolates. For this reason, it has been proposed that the hyper- variable regions may form surface-exposed loops in parasitized erythrocytes (Cheng et al., 1998, Kyes et al., 1999, Templeton, 2009a). Recent experimental data suggest that STEVOR and at least a subset of A-type RIFINs have only one TM segment near the C-terminus with the N-terminus (including the hypervariable loop region) exposed on the surface of the infected erythrocyte (Bultrini et al., 2009, Niang et al., 2009, Bachmann et al., 2015, Goel et al., 2015)*

1.5.3.1 The *rif* multigene family

Although repetitive sequence elements interspersed in the *P. falciparum* genome had been reported as early as 1988 (Weber, 1988), it was only in 1999 with the early releases of whole genome sequencing data from chromosome 3 that Kyes *et al.*, identified *rif* (*repetitive interspersed repeat*) as a large multigene family that potentially encoded hypervariable membrane proteins (Kyes et al., 1999). *The rif* genes form the largest multigene family in *P. falciparum* with 157 intact copies in the 3D7 laboratory isolate genome (Gardner et al., 2002, Joannin et al., 2008).

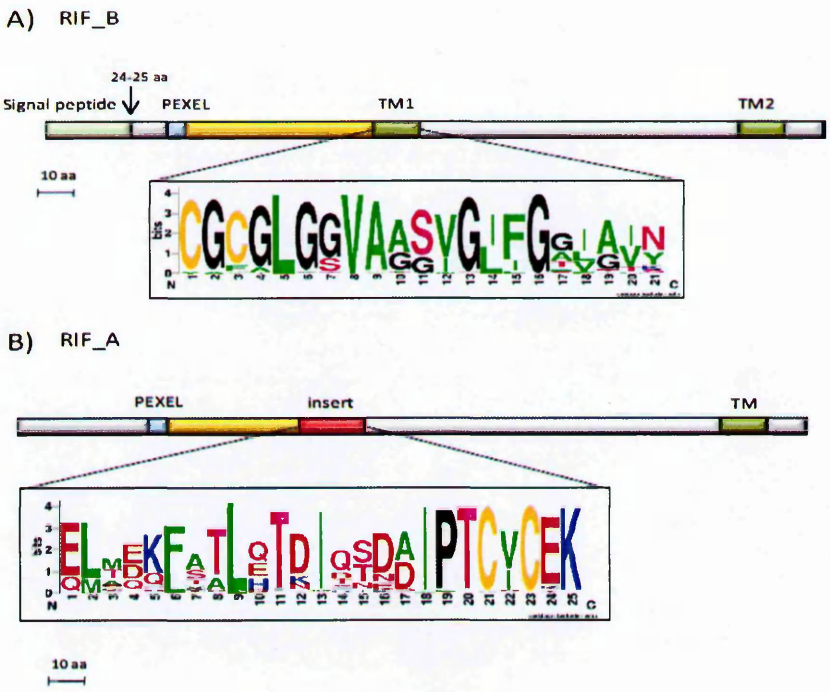


Figure 1.4. *Schematic diagrams of the predicted secondary structures of RIFINs. Diagram adapted from (Bultrini et al., 2009).*

The *rif* genes have been subdivided into two major groups *rifA* and *rifB* based on different sequence features (Figure 1.4.), most notable an insertion of a 25 amino acid consensus peptide, located about 66 amino acids downstream of the PEXEL motif in A-type RIFINs (Gardner et al., 2002, Joannin et al., 2008). The B-type RIFINs exhibit a greater degree of conservation than A-types and can be further sub-divided into at least 3 groups (Joannin et al., 2008, Wang et al., 2009) (Figure 1.5). Sequence analyses have shown that *rifB* genes have unique untranslated regions (UTRs) (Bultrini et al., 2009, Joannin et al., 2008), suggesting a functional significance of this sub-grouping. Importantly, despite the hypervariability observed within these multigene families, there are conserved groups and individual subsets of genes that are conserved across the sequenced laboratory isolate genomes (3D7, Dd2, IT and HB3) (Joannin et al., 2008, Wang et al., 2009, Bultrini et al., 2009, Claessens et al., 2011). A study by Claessens *et al.,* described a group of *rif* and *stevor* that are conserved across 4 laboratory isolates and Brazilian field isolates (Claessens et al., 2011). 13 *rif* and 5 *stevor* were shared in at least 3 of the 4 isolates with an overall amino acid similarity above 90% in a pair. In some instances the conserved *rif* and *stevor* occur next to each other on the chromosome. The intergenic region between such pairs is also conserved (Claessens et al., 2011). This observation suggests that there are conserved chromosomal regions in the genome that encode genes important for parasite survival. The strain-transcendent *rif* and *stevor* genes may themselves be functionally important or may be physically linked to genes important for parasite survival. It is therefore important to analyse *rif* and *stevor* diversity across phenotypically distinct field isolates. Some of the strain-transcendent *rif* and *stevor* are pseudogenes and it is possible that pseudogenes are more likely to be conserved because

they are not evolving under selection pressure from host immunity. These conserved pseudogenes might shed light on molecular evolution of these genes. Interestingly, the 3D7 genome also contains a distinct subset of 8 conserved *rifA* genes whose promoters are oriented head-to-head (i.e. share a common upstream (5') intervening region) with a Group A *var* gene (Lavstsen et al., 2003). Six of these group A *var*-linked *rif* have been classified into 2 different conserved sub-groups, named *rifA1* and those neighbouring the short Type 3 *var* genes, named *rifA3* (PFA0020w/PF3D7_0100400, PFD0025w/PF3D7_0400500, PFI1815c/PF3D7_0937500, PF11_0009/PF3D7_1100300, PF11_0520/PF3D7_1150300, PF13_0004/PF3D7_1300400) (Wang et al., 2009). Another unique *rifA* sequence, classified as *rifA2* (PF3D7_ 040600.1/PFD0070c), which has a unique 5' UTR, was also found to be conserved across parasite genomes (Wang et al., 2009).

1.5.3.2 The *stevor* multigene family.

The *stevor* gene family was first described as an expressed sequence, 7H8 detected by a monoclonal antibody (Limpaiboon et al., 1990). *stevor* is related to *rif* (Figure 1.5) in that they share some sequence homology at the 5' end and in the region encoding the cytoplasmic tail (Gardner et al., 1998), as well as having a similar gene architecture as shown in Figure 1.2. *stevor* forms a much smaller multigene family than the *rif* gene family, consisting of 31 intact genes in the 3D7 genome (Table 1.1). In comparison to *rif*, *stevor* also form a much more conserved gene family (Gardner et al., 1998, Cheng et al., 1998, Newbold, 1999). Phylogenetic analyses have classified *stevor* into two distinct groups per genome, a major group, which contains the majority of the sequences and a more diverse minor group (Blythe et al., 2009, Schreiber et al., 2008).

Conserved *stevor* genes have been observed across laboratory and field isolates (Albrecht et al., 2006, Lavazec et al., 2006, Blythe et al., 2009, Claessens et al., 2011)

There have been a limited number of studies that describe sequence diversity of *rif* (Joannin et al., 2008, Bultrini et al., 2009, Wang et al., 2009) and *stevor* (Lavazec et al., 2006) from several laboratory isolates with broad geographical origins and a few field isolates from Brazil (*rif* and *stevor*) (Albrecht et al., 2006), Gabon (*rif*) (Abdel-Latif et al., 2004), Kenya (*stevor*) (Blythe et al., 2009) and 5 isolates from West Africa sampled from traveller malaria patients (Bachmann et al., 2012). All studies describe a high divergence within the hypervariable loop regions of these proteins further supporting the hypothesis that host immune pressure drives antigenic diversification within this region (Lavazec et al., 2006, Templeton, 2009a). Full repertoires of *rif* and *stevor* repertoires from global collections of field isolates will enable detailed functional studies to establish whether they are targets of immune selection. Although whole genome sequence data is now available for hundreds of field isolates (http://www.broad.mit.edu, http://www.genedb.org/genedb/malaria/, http://www.sanger.ac.uk/Projects/P_falciparum) the data is fragmentary and inadequately annotated (Joannin et al., 2008, Wang et al., 2009, Lawton et al., 2012). Furthermore, optimized algorithms that allow automated assembly of the full *P. falciparum pir* gene repertoires are currently lacking and have thus far been based on algorithms that work best for *var* gene assembly (Newbold C., Otto, T, unpublished).

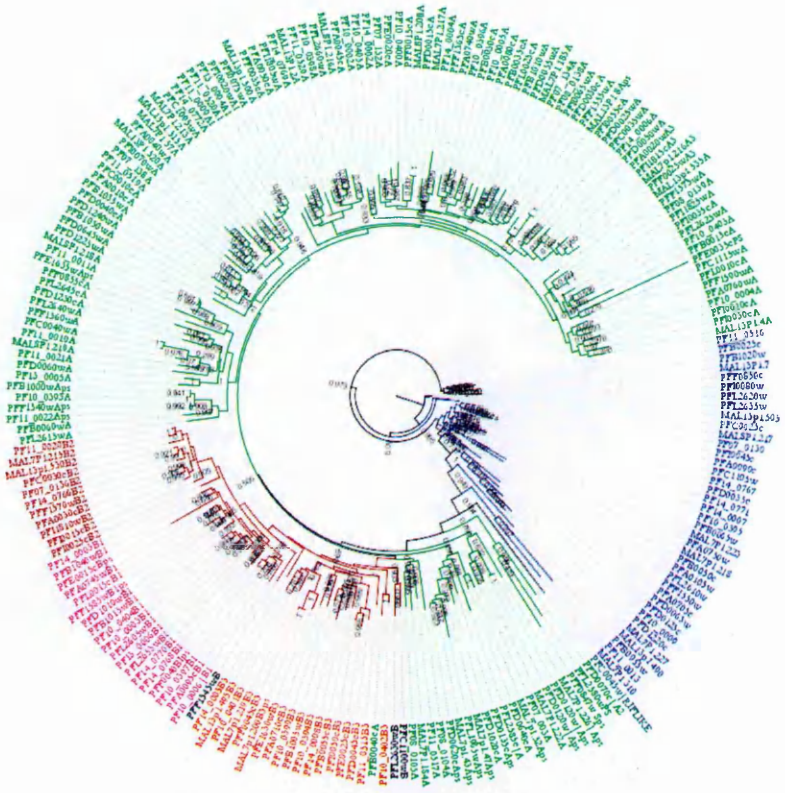


Figure 1.5. *Maximum likelihood tree of the pir multigene family from the 3D7 isolate showing clustering of stevor (blue) and rif genes into separate, but related, groups. The rif genes are divided into two major groups, A (green) and B, with B clustering further into B1 (pink), B2 (brown) and B3 (orange) subgroups (green). Three rif B sequences did not fall into either of the B subgroups (black). The tree was based on MUSCLE-alignments of 224 rif (n=184) and stevor (n=40) coding sequences, including pseudogenes or truncated genes in the 3D7 genome (version 2). These highly diverse and potentially recombining sequences resulted in trees with low statistical support and long branch lengths (Log-likelihood: - 140966). Tree constructed using PhyML (http://www.atgc-montpellier.fr), GTR model, viewed and edited using FigTree v1.4.0.*

1.5.4 Transcription of *rif* and *stevor* genes.

Microarray studies observed that gene expression in *P. falciparum* occurs in a tightly regulated cascade such that a gene is expressed just in time when the protein is required in the life cycle of the parasite (Bozdech et al., 2003). Early studies suggested VSA-encoding genes also followed a similar stage-specific expression profile, where each family is expressed sequentially, in overlapping windows, during the parasite's asexual intraerythrocytic developmental cycle. *var* transcription occurs in rings (Kyes et al., 2000, Llinas et al., 2006), followed by *rif* in late rings to early trophozoites (Kyes et al., 1999, Kyes et al., 2000, Llinas et al., 2006) and *stevor* in trophozoites (Kaviratne et al., 2002, Lavazec et al., 2007, Niang et al., 2009). However, subsequent studies showed within the *P. falciparum pir* gene family, different *rif* and *stevor* genes are expressed at multiple life cycle stages (Florens et al., 2002, Petter et al., 2008, Wang et al., 2010, Khattab et al., 2008, Mwakalinga et al., 2012, Tiburcio et al., 2012, Bachmann et al., 2012) as well as multiple stages during the IDC. *rif* transcripts have been observed throughout the intra-erythrocytic cycle (Bachmann et al., 2012, Wang et al., 2009), although peak expression of some members occurs at the late-ring to early-trophozoite (Kyes et al., 1999, Kyes et al., 2000), and of others in schizonts, merozoites, gametocytes, gametes and sporozoites (Petter et al., 2008, Mwakalinga et al., 2012, Wang et al., 2010). *stevor* transcripts have been observed in the ring stage (Bachmann et al., 2012), mid-trophozoite to late schizont stage (Niang et al., 2009, Kaviratne et al., 2002, Sanyal et al., 2012), merozoites (Khattab et al., 2008, Bachmann et al., 2012, Khattab and Meri, 2011) and in gametocytes (Sutherland, 2001, McRobert et al., 2004, Sharp et al., 2006, Tiburcio et al., 2012). Bachmann et al., looking at expression of VSA in laboratory isolates and five field

isolates from patients with traveller's malaria showed two distinct peaks in *rif* and *stevor* transcription in the field isolates, at ring and trophozoite stages, disputing the proposed cascade-like expression pattern for the expressed repertoires of *rif* and *stevor* genes as initially described (Bachmann et al., 2012). Collectively, these data suggest that the diverse sub-groupings or clades of the *rif* and *stevor* gene families (Janssen et al., 2004, Joannin et al., 2008, Wang et al., 2009, Bultrini et al., 2009, Blythe et al., 2009) might reflect the evolution of specialized functions for different sub-groups in different life cycle stages of the parasite (Otto et al., 2014a), in addition to their postulated roles in evading host immune responses. Laboratory isolates show low level of expression of *rif* and *stevor* compared with that of field isolates (Blythe et al., 2008, Bachmann et al., 2012, Daily et al., 2005), again hinting at important biological role(s) in host-parasite interactions.

There is evidence of transcriptional switching in the *rif* and *stevor* genes, though not in a mutually exclusive manner like the var genes (Lavazec et al., 2006, Lavazec et al., 2007, Cabral and Wunderlich, 2009). The strictly mutually exclusive expression of *var* genes (Scherf et al., 1998) is not observed for other multigene families. Multiple *rif* and *stevor* transcripts have been detected in bulk parasite populations as well as in populations of cloned lines *in vitro* (Bachmann et al., 2012, Kaviratne et al., 2002, Niang et al., 2009, Petter et al., 2008, Petter et al., 2007, Rovira-Graells et al., 2012). Transcription of multiple *stevor* in a single IE was clearly illustrated by Kaviratne and group where different *stevor* sequences were obtained from cloning and sequencing reverse transcription (RT) PCR products from individual trophozoites (Kaviratne et al., 2002). Although the processes involved in transcription regulation of VSA families may appear

to be very different, the factors controlling these processes may be shared. Evidence suggesting that transcription of VSA might be under shared control came from a study by Howitt et al., (Howitt et al., 2009), where parasites were transfected to carry a *var* promoter-driven blasticidin-S-deaminase (*bsd*) selectable gene cassette. By growing the parasites in increasing doses of blasticidin the investigators were able to increase the numbers of episomally replicating plasmids containing the respective promoters. High expression of the episomal *var* promoter not only led to repression of the endogenous expressed *var* gene, but additionally resulted in reduced expression of other VSA, *rif*, *stevor* and to a lesser extent *pfmc-2tm*. In the reciprocal experiments, transgenic parasites carrying *rif*, *stevor* or *pfmc-2tm* promoter-driven drug selectable constructs were used to examine the effects on endogenous *var* expression. Parasites expressing high copy numbers of *rif* or *stevor* promoters episomally had decreased *var* expression, while parasites overexpressing *pfmc-2tm* promoter-driven selectable markers showed no difference in *var* expression. Using this promoter titration strategy the authors demonstrated that VSA transcription requires a shared transcription factor. The overlap between transcriptional regulation of the individual variant gene families is still unclear, although the authors suggested that *var* and *rif* expression might be linked due to the overlap in expression timing of both gene families. The notable upregulation in *pfmc-2tm* genes in transfected parasites where *stevor* had been downregulated suggested potentially complementary roles for these two gene families in parasite survival.

Interestingly an earlier study observed co-regulation of *var* genes with *upsA* promoters (Group A *var*) and their neighbouring *rif* genes linked in a head-to-head direction (*rifA1*) (Tham et al., 2007, Wang et al., 2009). More recent work showed similar temporally

linked expression profiles between group *rifA1* genes and their linked *upsA-var* genes. In addition, activation of *rifA1* promoter resulted in knock down of endogenous *var*, a similar effect to that of *var* promoter (Witmer et al., 2012). This study demonstrated that the *rif* promoter had a direct effect on *var* gene expression levels. Additional evidence for linked transcription between *upsA var* genes and *rifA1* genes comes from analysis of *var* expression in parasites selected on human brain endothelial cells (HBEC) (Claessens et al., 2012). HBEC-selected parasites showed up-regulation of Group A *var* genes as well as the head-to-head-linked *rif* gene.

1.5.5 Localisation

P. falciparum parasites exist within a membrane-bound parasitophorous vacuole (PV) inside the IE. Blood stage *P. falciparum* extensively remodels its relatively inert host cell using a complex process of parasite-driven protein export involving translocation through several membranes (Maier et al., 2009). Exported proteins are trafficked through the endoplasmic reticulum, cross the parasite plasma membrane, through the PV membrane (PVM) and into the IE cytosol. Additionally membrane proteins including VSA have to cross through the IE cytosol to the IE membrane (Figure 1.6). Most exported proteins contain signal sequences that direct them into the secretory pathway. The signal peptide mediates entry into the endoplasmic reticulum and translocation across the parasite plasma membrane to the PV (Wickham et al., 2001). To cross the PVM and enter the IE, an additional signal, Plasmodium export element (PEXEL/VTS), consisting of a pentameric motif RxLxE/Q/D is required (Hiller et al., 2004, Marti et al., 2004). An aspartyl protease, plasmepsin V mediates cleavage of the PEXEL at the leucine residue (Boddey et al., 2013). Exported proteins which contain a PEXEL/VTS motif are translocated through the PVM into the IE via the translocon complex, PTEX (Plasmodium translocon of exported proteins) (de Koning-Ward et al., 2009, Riglar et al., 2013, Elsworth et al., 2014, Beck et al., 2014). Parasites also export a number of proteins that lack a PEXEL/VTS motif (PEXEL-negative exported proteins or PNEPs) using transport pathways that have not yet been fully characterized (Spielmann and Gilberger, 2010, Gruring et al., 2012).

Both RIFIN and STEVOR have a signal peptide as well as the PEXEL/VTS motif for export to the IE surface. PfEMP1 lacks a signal peptide and a classical PEXEL motif but

contains conserved motifs that function as PEXEL/VTS motif (Hiller et al., 2004, Marti et al., 2004). RIFIN, STEVOR and PfEMP1 are targeted for transport via the Maurer's clefts (MC), a network of parasite-derived membranous structures that first appear in the early ring stages of parasite development within the IE (Haeggstrom et al., 2004, Hiller et al., 2004, Marti et al., 2004, Haeggstrom et al., 2007, Gruring et al., 2012). Trafficking of STEVOR and PfEMP1 to the IE membrane via the MC are both dependent upon a protein called *P. falciparum* PfEMP1 trafficking protein one (PfPTP1) (Maier et al., 2008, Rug et al., 2014).

The A- and B-RIFIN sub-groups are reported to have distinct subcellular localizations within the IE. A-RIFINs are translocated into the MC and are associated with the IE cell surface while B-RIFINs reside within the parasitophorous vacuole (Petter et al., 2007, Petter et al., 2008, Bachmann et al., 2012). Surface iodination and immunoprecipitation experiments suggested localization of RIFIN at the IE surface (Fernandez et al., 1999, Kyes et al., 1999). In merozoites, the A-RIFINs are localized at the apical tips while the B-RIFINs are found inside the merozoite cytosol (Petter et al., 2007). More recently a B-RIFIN was found to be preferentially expressed in merozoites, gametocytes and gametes, and localized to the surface of merozoites, gametes and sporozoites (Florens et al., 2002, Petter et al., 2008, Wang et al., 2010, Mwakalinga et al., 2012).

Studies based on immunofluorescence assays have located STEVOR within the MC in trophozoites and closely apposed to the IE surface in late trophozoites (Lavazec et al., 2006, Blythe et al., 2008, Kaviratne et al., 2002). Later a study using a live immunofluorescence assay showed that STEVOR are located on the surface of IE (Niang et al., 2009). STEVOR have also been shown to translocate to the apical ends of

merozoites (Blythe et al., 2008), and to the surface of invading merozoites (Khattab et al., 2008, Khattab and Meri, 2011). STEVOR proteins have also been detected in gametocytes (McRobert et al., 2004, Tiburcio et al., 2012) and in sporozoites (Florens et al., 2002).

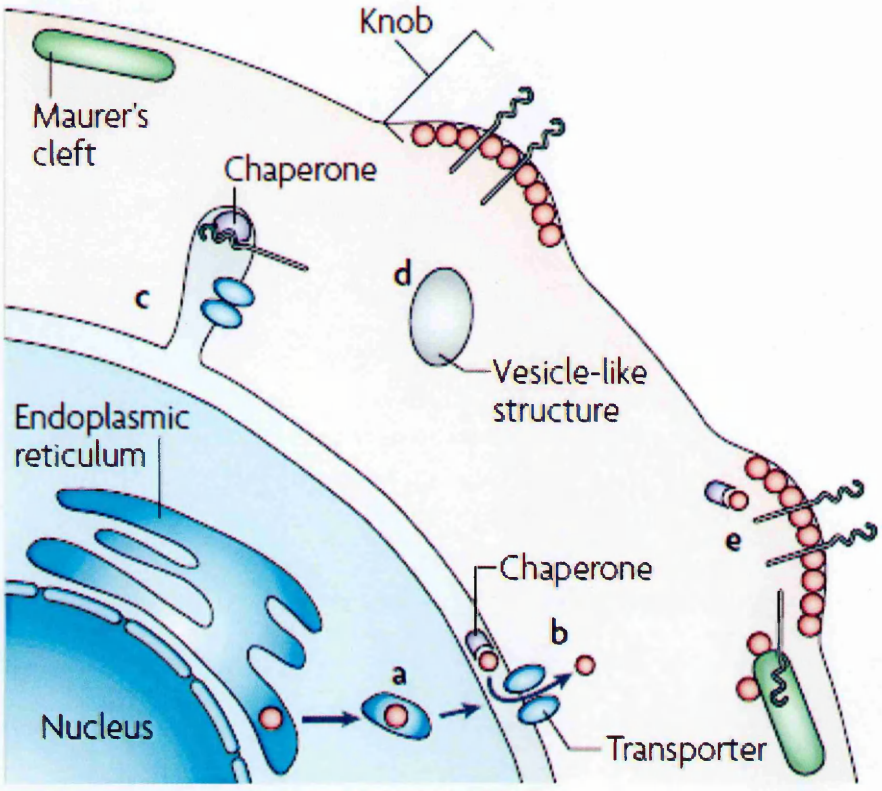


Figure 1.6. *Many-exported proteins in P. falciparum infected erythrocytes (IE) contain a PEXEL/VTS motif that is cleaved in the ER and targets the protein into the red blood cell. Some IE surface membrane-associated parasite proteins are thought to be transported by the translocon complex PTEX across the PVM into MC which carries the protein to the IE surface. Adapted from Maier et al 2009.*

1.5.6 RIFIN and STEVOR function.

RIFINs were initially thought to be involved in rosetting and were termed rosettins (Helmby et al., 1993). Subsequent studies showed that PfEMP1 mediated this phenotype (Rowe et al., 1997), though whether or not RIFINs play a complementary role in rosetting remained unclear. A structural analysis has suggested different architectures of the different RIFIN sub-groups, where B-RIFINs are predicted to have a signal peptide and two TM domains, while a majority of the A-RIFINs do not have a conventional signal peptide and *in silico* and experimental data suggest the presence of only one TM domain (Bultrini et al., 2009, Niang et al., 2009, Goel et al., 2015). A similar topology is predicted for *stevor* genes (Niang et al., 2009). Additionally a study on a splenectomized malaria patient revealed lack of expression of *rifA* as well as *var* and *stevor* while *rifB* and *Pfmc-2TM* were expressed in the infecting parasite that also lacked cytoadherent properties (Bachmann et al., 2009) suggesting different functional properties of the two major *rif* groups. A recent study that showed A-RIFINs mediate rosetting of blood group A erythrocytes (Goel et al., 2015). The study first demonstrated the ability of RIFINs to bind RBC using A-RIFIN-transfected Chinese hamster ovary (CHO) cells. The transfected CHO cells expressed RIFIN on the surface and preferentially bound blood group A RBCs via blood group A antigens to form large rosettes. Treatment of blood group A RBCs with neuraminidase to convert them into group O RBCs by cleaving off the α3 N-acetylgalactosamine residue, reduced binding to the RIFIN-expressing CHO cells. The authors then showed that a specific *rif* together with a *var* gene were highly expressed in the rosetting parasite line FCR3S1.2 when the parasite was enriched for rosetting. Anti-RIFIN IgG disrupted rosettes in this parasite when cultured in blood group

A but did not affect rosetting when cultured in blood group O. This supported the role of the specific FCR3S1.2 RIFIN in rosetting of FCR3S1.2-IE with group A RBCs. Transfecting a low-rosetting parasite line with this specific RIFIN increased the ability of the parasite to form rosettes, and more so with blood group A RBCs. These results demonstrated that A-RIFINs could play a complementary role to PfEMP1 in rosetting. STEVOR have been shown to undergo clonal variation at the IE surface (Lavazec et al., 2007, Blythe et al., 2008, Niang et al., 2009) and thus may play a role in immune evasion of the IE along with simultaneously expressed RIFIN and PfEMP1. However, besides a potential role in immune evasion, STEVOR, like RIFIN, may serve specialized roles in the life cycle of the parasite. Recent studies imply a role for STEVOR in malaria pathogenesis. In a study looking at changes in IE deformability, STEVOR overexpression increased the rigidity of the trophozoite-IE (Sanyal et al., 2012) and gametocyte-IE (Tiburcio et al., 2012). Infected and uninfected erythrocytes were perfused through a defined matrix of microspheres with flow and retention rates determined. Deformability of IE was also measured using ektacytometry via Laser-assisted optical rotational cell analyzer (LORCA). Increased deformability index as measured by ektacytometry analysis as well as increased retention rates were observed in parasites that overexpressed *stevor* compared to parasites that did not (Sanyal et al., 2012, Tiburcio et al., 2012). Synthetic STEVOR peptides can bind to human erythrocytes (Garcia et al., 2005), which supports a possible role in rosetting. Localization of STEVOR at the apical ends of merozoites (Blythe et al., 2008), and on the surface of invading merozoites (Khattab et al., 2008, Khattab and Meri, 2011), suggests a role in invasion, or immune evasion by the merozoites. More recent data has confirmed that STEVOR antigens can also mediate

PfEMP1-independent rosetting of IE and may play a role in invasion via the glycophorin C pathway (Niang et al., 2014). This study first demonstrated that STEVOR proteins were co-localised on live merozoite surface with the merozoite surface protein 1 (MSP1), the well characterised merozoite surface antigen key in the invasion process. Anti-STEVOR sera raised against the semiconserved regions of 2 STEVOR antigens were able to inhibit invasion. The study then demonstrated that STEVOR mediated binding to RBCs via Glycophorin C, and that this rosetting in turn enhanced invasion.

VSA are non-essential for parasite survival *in vitro.* VSA gene knock out experiments (Dzikowski and Deitsch, 2008, Howitt et al., 2009) and a study looking at *stevor* expression in laboratory and field isolates (Blythe et al., 2008) showed that parasite growth was not affected by downregulation of these genes. A different study looking at malaria in a splenectomized individual showed that the parasites causing the infection did not express *var*, *rifA* and *stevor* (Bachmann et al., 2009). As mentioned above, different groups of RIFIN and STEVOR are expressed at the host-parasite interface, i.e. the IE surface, merozoite surface, mature gametocytes, sporozoites and gametes, suggesting they have important functions in parasite biology (Florens et al., 2004, McRobert et al., 2004, Petter et al., 2008, Niang et al., 2009, Khattab and Meri, 2011, Mwakalinga et al., 2012, Tiburcio et al., 2012, Goel et al., 2015, Niang et al., 2014). In the rodent malaria parasite *P chabaudi*, a difference in *cir* gene expression was observed between mosquito-transmitted parasites and serially blood-passaged parasites. The switch in *cir* expression was linked to a change in virulence (as measured by parasite multiplication rates), with fewer *cir* genes expressed and increased virulence in the serially blood passaged parasites (Spence et al., 2013). Taken together, these studies suggest important roles for *pir* genes

during host-parasite interactions, specifically immune evasion strategies and/or functional roles in infection and virulence.

1.5.7 Immune responses to RIFIN and STEVOR.

Few studies have described immune responses to RIFIN and STEVOR. A cross sectional study in a hyperendemic area in Gabon reported higher recognition of recombinant RIFIN in exposed immune adults as compared to children, although antibodies were also detected in the children (Abdel-Latif et al., 2002). A second study showed that high levels of antibodies to RIFIN were associated with rapid clearance of parasites in children (Abdel-Latif et al., 2003). There is evidence that responses to RIFIN, in addition to PfEMP1, can contribute to the overall anti-VSA antibody response formed against the IE surface (Abdel-Latif et al., 2004). In this study immune sera from Gambian adults was preadsorbed with recombinant RIFIN, which resulted in a reduction in the reactivity to the IE surface. In another study, patients with severe malaria in Ghana had higher antibody levels to RIFIN than asymptomatic controls suggesting that antibodies to RIFIN are boosted by current infections (Schreiber et al., 2006).

Even fewer studies have looked at the role of STEVOR as targets of immune responses. Sera from immune adults recognized recombinant STEVOR proteins (Schreiber 2008). In addition a longitudinal study in 9-month old infants showed that although there was no correlation with protection, antibody levels to a recombinant STEVOR were higher in children that experienced a higher frequency of malaria episodes (Schreiber et al., 2008). Recently, Chan et al., have shown that PfEMP1 is the major target of anti-VSA antibodies targeted to IE. Genetically modified parasites with reduced expression of

PfEMP1 were produced by transfecting parasites with a construct containing *var* promoter without the downstream *var* gene. Drug driven expression of the var promoter resulted in silencing of endogenous *var* genes. Sera from immune Kenyan adults showed decreased surface binding to PfEMP1 knockdown parasites in comparison to wild type parasites suggesting that PfEMP1 forms the major target of IE surface antibody response (Chan et al., 2012). However residual reactivity to the PfEMP1 knockdown parasites was observed suggesting that RIFIN and STEVOR might still serve as immune targets on the IE surface (Voss et al., 2006, Chan et al., 2012).

More comprehensive studies are required to establish the functional role of these PIR antigens in development of protective immunity against malaria, as well as in parasite biology and disease pathogenesis, and how they interact with the most well characterized virulence factors in *P. falciparum*, PfEMP1.

1.6 Thesis objectives.

1.6.1 General objective

To assess whether the variant antigens encoded by the *rif* and *stevor* multigene families contribute to the pathogenic and antigenic properties of *P. falciparum*-infected red cells.

1.6.2 Specific objectives

1) To characterise the sequence diversity of *rif* and *stevor* in culture-adapted and clinical isolates of *P. falciparum* from different geographical locations.

2) To profile *rif* and *stevor* gene expression patterns during asexual blood stage development of parasite isolates exposed to different selection pressures.

3) To investigate whether the hypervariable loop region of RIFIN antigens are targets of naturally acquired antibodies.

CHAPTER 2

2. Materials and Methods

2.1. Study area and study participants

The work discussed in this thesis was carried out at the KEMRI-Wellcome Trust Research Programme located in Kilifi County along the Kenyan coast, north of Mombasa. The Programme works in partnership with the Kilifi County Hospital. The hospital serves the entire Kilifi population of about 544,000. Kilifi is a malaria endemic area, typically experiencing two seasons of high malaria transmission, May to June and December to January following the long and short rains respectively. Malaria transmission has been on the decline from 1999 (Okiro et al., 2007, Ceesay et al., 2008, Kleinschmidt et al., 2009, O'Meara et al., 2008) but hotspots of both febrile and asymptomatic malaria still exist (Bejon et al., 2010, Bejon et al., 2014). The major vector has been *Anopheles gambiae* sensu lato (s.l.) with *Anopheles funestus* playing a minor role (Mbogo et al., 2003), but recent studies have observed vector replacement along the Kenyan coast with *Anopheles arabiensis* and *Anopheles merus* becoming major vectors (Mwangangi et al., 2013). Chloroquine was used as the first-line treatment for non-complicated malaria until 1998 when the switch was made to sulphadoxine-pyrimethamine (Shretta et al., 2000). Artemether-lumefantrine replaced sulphadoxine-pyrimethamine in 2006 (Amin et al., 2007).

2.2. Sample collection and processing

Isolates used in this study were sampled from children admitted at the Kilifi County hospital with either severe malaria (part of an ongoing *var* study led by Dr. P. Bull) or *ex*

vivo samples collected from children with non-severe malaria (part of an ongoing microarray study led by Dr. M. Mackinnon). Additional isolates were collected from children with asymptomatic malaria during annual cross-sectional surveys for malaria in 2 field cohorts, RTS,S cohort (2007-2012) (Bejon et al., 2006) and Junju cohort (2006-2015) (Mwangi et al., 2005) in Kilifi County. 2-5 ml of blood was collected from each child and 2-4.4 ml transferred into heparinized vacutainers containing 50units of heparin. The samples were transported to the laboratory on ice.

In the laboratory, the samples were centrifuged at 2000 revolutions per minute (rpm) for 5 minutes to separate plasma from the pellet. Plasma was harvested into 2ml cryotubes and stored at -80^{0}C. The remaining pellet was resuspended in 5ml "yellow RPMI" (RPMI medium pH 7.2 containing 1mM L-Glutamine, 25ug/ml Gentamicin and 200mg D-glucose/ml). To separate the buffy coat and peripheral blood mononuclear cells (PBMCs) from the RBC pellet, the resuspended pellet was layered onto an equal volume of Lymphoprep ® (Axis-shield poc AS cat. No. 1114545, Norway), and centrifuged at 2000 rpm for 20 minutes, without brakes. The PBMCs were harvested and stored at -80^{0}C. The remaining pellet was resuspended in equal volume yellow RPMI. To remove granulocytes from the resuspended pellet, a suspension of the pellet and 0.6 x volume Plasmion ® (Bellon France cat. No. 00-13-6676) was made and incubated at 37^{0}C for 10 minutes. The supernatant containing granulocytes was harvested, washed in yellow RPMI and stored at -80^{0}C. The RBC pellet was resuspended in yellow RPMI to a final volume of 5ml and split into 3 aliquots for storage of ring-infected RBCs in glycerolyte (parasite stabilate), in Trizol™ (Life Technologies, Invitrogen) for RNA extraction, (ring RNA, acute stage) and the rest for culturing to trophozoite stage (approximately 24 hours). The

pellet for culturing was resuspended in 5ml "blue RPMI" (yellow RPMI with 10% AB serum from adult non-immune European donors) and cultured for 24 hours or until the parasites matured to pigmented trophozoite stage as determined by Giemsa staining. The cultures were split into two aliquots for storage of trophozoite-infected RBCs in glycerolyte, and in Trizol™ (trophozoite RNA, acute stage). All the aliquots for RNA preparation were frozen in Trizol™ at -80^{0}C. The ring and trophozoite parasite stocks in glycerolyte were stored in liquid nitrogen at -180^{0}C.

2.3. Parasite culture methods

2.3.1. Thawing cryopreserved *P. falciparum* and culture establishment

Cryopreserved ring-infected erythrocytes were thawed through gradual addition of decreasing concentrations of saline solution to restore parasites to isotonic condition. The frozen stocks were retrieved from liquid nitrogen and rapidly thawed in 37^{0}C water bath for 1 minute. The IE were transferred to a 50ml falcon tube (Greiner Bio-One). While gently shaking the falcon tube, sodium chloride (NaCl) was added drop-wise to the IE in order of decreasing concentration. First 200µl of 12% NaCl was added followed by 10ml of 1.8% NaCl and 10ml of 0.9% NaCl. The IE suspension was centrifuged at 1800rpm for 5 minutes and the supernatant removed by aspiration. The IE pellet was then washed twice by adding 10ml yellow RPMI and centrifuging at 1800rpm for 5 minutes. After the final wash the pellet was resuspended in 5ml blue RPMI (yellow RPMI with 10% pooled serum from malaria non-exposed non-immune Kenyan donors or 10% AB serum from non-immune European adults) and transferred to a 50ml flask for continuous culture. Two aliquots of approximately 100µl were taken for parasite assessment by Giemsa staining of blood films, and for mycoplasma testing.

Established parasite cultures were grown in complete medium (RPMI medium pH 7.2 containing 1mM L-Glutamine, 25μg/ml Gentamicin and 200mg D-glucose/ml and 10% pooled serum), and in a gas mixture of 3% O_2, 5% CO_2 and 92% N_2 at 37^0C. The cultures were maintained at 10% RBC hematocrit and below 10% parasitemia.

2.3.2. Assessing parasite growth using Giemsa-stained blood films

Thin blood films were prepared from cultures by applying a drop of the culture pellet onto glass slides (Thermo Fisher Scientific Inc.) and spreading the blood very thinly using a slide. The smears were air dried and fixed using methanol (Analar ® VWR International). The smears were stained with 10% Giemsa stain (Sigma Aldrich) for 10 minutes after which they were rinsed in water. The slides were left to air dry and examined under a light microscope (Leica Microsystems) at 1000x magnification in oil immersion.

Parasitemia was determined by calculating the percent of IE per 1000 intact erythrocytes at random fields on the slide. Cultures with parasitemia of 10% and above were diluted by adding O^+ uninfected erythrocytes.

2.3.3. *P. falciparum* synchronisation

Parasites were regularly synchronized using 5% D-sorbitol (Sigma Aldrich), which causes selective lysis of late-stage parasite-IE. Cultures with predominantly ring-IE were transferred to 15ml falcon tubes (Greiner Bio-one), centrifuged at 2000 rpm for 5 minutes and the supernatant aspirated. 10x pellet volume of pre-warmed 5% D-sorbitol was added and the mixture incubated in a water bath at 37^0C for 10 minutes. The cultures were then washed three times with yellow RPMI and centrifuged at 1800 rpm. Assessment of synchronization success was done by examination of Giemsa-stained slides prepared

immediately after synchronization, with a successful synchronization resulting in cultures with predominantly ring stage parasites.

2.3.4. Mycoplasma testing and treatment

Parasite cultures were regularly tested for mycoplasma contamination and immediately treated or discarded if found to be infected (Rowe et al., 1998). Parasite DNA was extracted from 100µl of culture by freeze-thaw method or using the QIAamp DNA Blood Mini Kit (Qiagen). A nested polymerase chain reaction (PCR) targeting 16s and 23s rRNA spacer regions on *Mycoplasma* and *Acholeplasma* was used to detect mycoplasma contamination (Tang et al., 2000).

Freshly thawed parasites from frozen stocks and new O^+ uninfected erythrocytes from donors were tested for mycoplasma after one overnight culture. Cultures that tested positive for *Mycoplasma* or *Acholeplasma* were put through a 7-day treatment regime using Mycoplasma Removal Agent (AbD Serotec, Bio-Rad), and re-tested to ensure the contamination had been eliminated.

2.4. Time course experiments

Time course experiments were conducted to analyse *rif* and *stevor* expression patterns at different stages of asexual parasite development within red blood cells and to establish the peak expression points for the two gene families (see Appendix Table 9.5 and 9.6). For the laboratory isolates 3D7, IT and HB3, RNA was extracted from synchronized cultures at specific time points: 0 hours post invasion (hpi), 12 hpi, 18hpi, 24hpi, 30hpi and 40hpi corresponding to early ring, late ring/ early trophozoite, mid trophozoite, late trophozoite and schizont stages, respectively. For the culture adapted field isolates RNA

was extracted at four time points across the IDC, early to mid ring, late ring to early trophozoite, mid to late trophozoites and schizont stages as determined by examining Giemsa-stained blood films. For the *ex vivo*-cultured, wild (non-culture-adapted) Kilifi isolates, RNA was extracted at 7 time points every 10 hours across a single IDC. For the culture adapted laboratory and field isolates, to ensure that synchrony was maintained throughout the time course, two sorbitol synchronizations, eight hours apart were done at beginning of each time course experiment (see Figure 5.1. and Appendix Tables 9.5 and 4.6). Parasites were harvested from culture into Trizol™ for RNA extraction (Kyes et al., 2000).

2.5. Extraction of *P. falciparum* RNA and genomic DNA (gDNA)

2.5.1. RNA extraction from *P. falciparum* parasites

To harvest RNA 100µl parasite pellets were prepared by spinning the appropriate volume of cultures at 2000 rpm for 5 minutes and aspirating the supernatant. To cultures containing predominantly ring-IE 10x pellet volume of Trizol™ was added and to cultures containing mainly trophozoite-IE, 20x pellet volume Trizol™ was added. The Trizol™ suspensions were stored at -80^{0}C or immediately used for RNA extraction. RNA was extracted according to standard protocol (Kyes et al., 2000). 200µl of chloroform (Sigma- Aldrich) was added to 1000µl of Trizol™ sample, mixed vigorously and centrifuged at 4200 rpm at 4^{0}C for 35 minutes. The aqueous phase containing total RNA was harvested and mixed with Isopropanol (Sigma-Aldrich) and incubated overnight at 4^{0}C to precipitate the RNA. The RNA pellet was washed in ice cold 75% ethanol and

resuspended in 20µl RNA Secure™ (Life Technologies, Invitrogen) by heating at 60^{0}C for 10 minutes.

2.5.2. gDNA extraction from *P. falciparum* parasites

To obtain parasite gDNA, infected erythrocytes were lysed by adding 0.2% Saponin (Sigma-Aldrich) in 1x PBS to culture pellets and vortexing briefly. The parasite DNA pellet was harvested by spinning the culture-saponin mixture at 5000 rpm for 10 minutes. The DNA pellets were washed twice in 1x PBS and resuspended in 200µl PBS. Samples in PBS were used for DNA extraction using the QIAamp DNA mini blood kit (Qiagen) according to manufacturer's instructions.

2.6. Generating gDNA and cDNA libraries

2.6.1. cDNA preparation

Prior to cDNA preparation, gDNA was removed from the total RNA preparation (**Section 2.5.1.**) by treatment with DNase. 1-2 units of DNase (Ambion) and 20µl final volume DNase buffer was added to total RNA and the mixture incubated at 37^{0}C for 1 hour. First strand cDNA was synthesized using the SuperScript® III kit (Invitrogen) and random hexamers following manufacturer's instructions. For every RNA sample, two reactions were set up. In one reaction 1 unit of SuperScript®III, the reverse transcriptase enzyme, was added (RT+), while in the control reaction the enzyme was omitted (RT-). The cDNA was stored in aliquots at -20^{0}C.

2.6.2. Reverse transcriptase PCR (RT-PCR)

RT-PCR was used to analyse *rif* and *stevor* expression patterns in the laboratory and wild isolates. The different primer combinations used to amplify *rif* and *stevor* fragments are listed in Table 2.1.

For *rif* expression analysis a set of published primers (Kyes et al., 1999) with slight modifications were used. For the two different amplicon sequencing approaches, capillary and 454 sequencing, different *rif* reverse primers were used. The original reverse primer was used (Kyes et al., 1999) for most isolates where amplicons were first cloned and then sequenced using the capillary method, whereas a modified version was used for all the amplicons sequenced using clone-free 454 (Table 2.1).

For *stevor* gene sequencing, different sets of primers were used in the different sequencing approaches. For amplicon cloning followed by capillary sequencing, primers published by Lavazec *et al.* were used (Lavazec et al., 2006), while for 454 amplicon sequencing a combination of a new *stevor* degenerate forward primer designed using Primer3 (http://primer3.sourceforge.net) (Stevor-primer3_F_72), and a published reverse primer (Albrecht et al., 2006) was used. For each PCR, the component volumes and concentrations were optimized by titrating dNTPs (0.2-0.4mM), primers (0.4-1mM) and Taq polymerase (0.5-1 units).

Cycling conditions were optimized by setting up different PCR reactions to run in parallel on a gradient thermocyler at different annealing temperatures. A gradient PCR was set up for every new primer pair and and/or polymerase. For most amplicons for cloning and/or sequencing the high fidelity Platinum Taq Polymerase High Fidelity enzyme was used.

For a subset of amplicons for 454 sequencing the regular Platinum Taq Polymerase, with 6 X less fidelity was used.

For each *rif/stevor* PCR, gDNA was included, (1) as a positive control to ensure the PCR conditions and primers were optimal for amplification from cDNA, (2) to sample the genomic repertoire matched for every isolate used to assay *rif/stevor* expression, and (3) to determine whether there was any bias in the sampling of transcripts from cDNA.

Table 2.1. *Degenerate primer sets used to generate rif and stevor amplicons for sequencing.*

Primer Name	Sequence (space after each set of 3)	Length (bases)
rif		
[a]Kyes-rifF1	C[A/G]T CAC GA[G/T] TGT TAA GCG	18
[a]Kyes-rifF2	CGA [A/G][C/T]G TGA ATT GTA TGC	18
[a]Kyes-rifF3	C[C/T]A C[C/T]A G[A/G]T TAT TAT GCG	18
[a]Kyes-rifRev #	CTT CA[A/T] ATT [A/G]TT [A/T]TT T[C/T][G/T] [A/G/T]CG ATA ACG	
[a]Kyes-rifR1-VT@	CTT CA[A/T] TTT [C/T]TT [A/T]TT T[C/T][G/T] [G/A/T]C[G/T] ATA ACG	27
stevor		
[b]Lavazec_L	CAA AAG GAA GAG ATA AGT AT	20
[b]Lavazec_R	GTT TCT TGC ATT CAT GTT TCC	21
[c]Stevor-primer3_F_72	[A/T][A/G/T][A/C/T] AGA ACC [A/C/T/G]AT GTC AAC G	19
[d]Albrecht_R	CAT AAA TGT TTC TTG CA[C/T] TCA TG	23

a: Complex rif primer set with 3 forward and 1 reverse primer ((Kyes et al., 1999), with the original and modified reverse primer. Changes highlighted in red text.

#: Original rif reverse primer from publication.

@: Modified rif reverse primer

b: stevor primers set 1 (Lavazec et al., 2006) used for cloning/capillary sequencing

c: Newly designed (using Primer3) stevor forward primer for stevor primer set 2 (Thathy et al., unpublished)

d: Published stevor reverse primer for stevor primer set 2 (Albrecht et al., 2006). Primer set 2 used to amplify stevor for 454 sequencing

2.6.3. Gel electrophoresis of amplicons

Agarose gel electrophoresis was used to analyse amplicons and to purify amplicons for cloning and sequencing, as needed. 1% - 2% agarose gels were prepared by dissolving appropriate volumes of UltraPure™ Agarose (Invitrogen) in 1x Tris Acetate ethylenediaminetetraaceticacid (EDTA) (TAE) Buffer (Sigma-Aldrich) with 0.5mg/ml ethidium bromide (EtBr). Each sample was mixed with 5x (Qiagen) or 6x (Fermentas) loading dye before electrophoresis. 1 Kb Plus DNA Ladder™ (Invitrogen) was used as the DNA size marker in each gel for estimation of amplicon size. Gels were run in 1X TAE at 100-200 V depending the size of the fragments, agarose concentration and on the size of the gel tank used. A gel imaging system (Biorad) was used to visualize gels under UV light and image and analyze separated nucleic acids.

2.6.4. Amplicon purification and quantification

PCR fragments were purified using the QIAquick PCR Purification Kit (QIAgen) following the manufacturer's instructions with slight modifications, including elution with 1X TE in place of elution buffer and reduced volume of elution solution, (using between 6 to 35µl 1X TE). An additional wash with 35% guanidine hydrochrolide aqueous solution prior to elution was included to remove primer dimers. For a subset of amplicons with large amounts of primer dimers a prior purification through sephacryl columns was done prior to purification using the QIAquick PCR Purification Kit. In addition a subset of *rif* amplicons were purified by gel extraction of bands using the QIAquick Gel Extraction Kit (QIAgen).

Purified amplicons were quantified using either the NanoDrop (1000 or 8000, Thermo Scientific) or Qubit® 2.0 Fluorometer (Life Technologies, Invitrogen).

2.6.5. TOPO-cloning

Purified amplicons were cloned into pcr2.1TOPO vector (Life Technologies, Invitrogen) following manufacturer's instructions. Briefly 2ml of amplicon was mixed with 1µl pcr2.1TOPO vector, 1µl salt solution (1.2 M NaCl and 0.06 M $MgCl_2$) and 2µl of water. The reaction was incubated at room temperature for 15-30 minutes and then placed on ice. 2µl of the reaction was added to 50µl of One Shot® Top10 competent *Escherichia coli* cells (Life Technologies, Invitrogen) and incubated on ice for 30 minutes. Transformation of *E. coli* was achieved by heat shock treatment by incubating the plasmid - *E. coli* mixture in a water bath at 42^0C for 30 seconds and immediately placing the tube on ice for 3 minutes. 250µl of SOC medium was added to the cells and incubated at 37^0C for 1 hour with agitation (200rpm). 50-200µl of the culture was plated on LB agar plates containing 50mg/ml ampicillin and 40mg/ml X-gal for blue white colony screening. Plates were incubated at 37^0C overnight. White colonies (i.e. transformed *E. coli* harboring plasmid vectors containing inserts) were picked for sub-culturing and plasmid DNA preparation. 20 clones from cDNA (per timepoint, and in certain cases for *rif* amplicons per upper and lower band per timepoint per sample) and 50 clones from gDNA were picked for screening and sequencing.

2.6.6. Plasmid DNA preparation

White colonies were picked and incubated in LB broth containing 50mg/ml ampicillin at 37^0C overnight with agitation (250rpm). Plasmid DNA was harvested using QIAprep Miniprep kit (Qiagen) following manufacturer's instructions. Briefly 1.5ml of the cultures were pelleted by centrifugation at maximum speed for 2 minutes. The bacterial pellet was dissolved in resuspension buffer (P1) containing RNase A. The cells were

lysed using alkaline lysis buffer (P2) and subsequently neutralized and adjusted to high-salt binding conditions for purification using silica membrane filters on microspin tubes. Plasmid DNA was analysed by agarose gel electrophoresis and screened for presence of intact inserts using restriction digestion.

2.7. Capillary sequencing

TOPO clones containing inserts (*rif* or *stevor* amplicon) as per restriction digestion results were sequenced using the BigDye terminator kit version 3.1 (Life Technologies, Invitrogen) and resolved with a 3130xl ABI (Applied Biosystems) capillary sequencer. Some of the plasmids were sent either to the Sequencing, Genotyping, OligoSynthesis and Proteomics (SEGOLIP) Unit at BecA-ILRI Hub (Nairobi, Kenya), or to Beijing Genomics Institute (BGI), Hong Kong, for capillary sequencing. For each plasmid, the insert was sequenced in both the forward and reverse directions using M13 forward and reverse primers (Appendix Table 9.7).

2.8. 454 sequencing

454 library preparation and sequencing of *rif* and *stevor* amplicons was performed at the Wellcome Trust Sanger Institute, Cambridge, UK. Purified amplicons were quantified using Qubit® 2.0 Fluorometer or PicoGreen® dsDNA quantitation assay (Life Technologies, Invitrogen). A starting concentration of 100ng of amplicon was used. For amplicons with concentrations less than 100ng, carrier amplicon of the same length but from an unrelated species (bovine) was added to make up a final concentration of 100ng. Amplicons were mixed with 1x TE to a final volume of 16µl.

Library preparation was done using the GS FLX Titanium Rapid Library Preparation Kit (Roche) using the Ligated Adaptors design following manufacturer's instructions. In

summary, the amplicons (bearing TA overhangs) were blunt ended by adding the fragment end repair mix containing polynucleotide kinase (PNK), PNK buffer, ATP, dNTPs, T4 polymerase and Taq polymerase and incubating at 25^0C for 20 minutes, 72^0C for 20 minutes and 4^0C hold. Rapid Library (RL) Multiplex Identifiers (MID) Adaptors from the GS FLX Titanium Rapid Library MID Adaptors Kit were ligated to the blunt ends of the amplicons by adding the RL MID adapters and RL ligase to the amplicons and incubating at 25^0C for 10 minutes. For our experiment, 48 RL MID adapters were used to allow multiplexing 48 different amplicons in a single sequencing pool. Small fragments were removed from the libraries by carrying out two rounds of solid phase reversible immobilization (SPRI) using Agencourt AMPure beads (Beckman Coulter) purification, as follows. The libraries were mixed with the SPRI beads to a 1:1.8 volumetric ratio in 1.5ml DNA LoBind tubes (Eppendorf) and incubated for 5 minutes at room temperature. The tubes were placed on a magnetic particle concentrator (MPC) till the beads were fully pelleted on the walls of the tubes. The supernatant was removed and discarded and the beads washed with 70% ethanol and resuspended in 1x TE. After the second round of purification the libraries were eluted in 50µl of 1x TE and quantified using PicoGreen®. The libraries were further quantified using quantitative real time PCR (QRT-PCR). Starting with the quantities obtained by PicoGreen® assay, the libraries were diluted to a starting concentration of 10^8 molecules in 1ml of 1x TE. Equimolar concentrations of 48 libraries were pooled based on QRT-PCR results.

Amplification of libraries was carried out through emulsion PCR using Lib-L emPCR Kit. Sequencing was done on the GS FLX+ (Roche).

2.9. Sequence assembly and analysis

2.9.1. Capillary sequence analysis.

For reads obtained by capillary sequencing, assembly and analysis were done using BioEdit Sequence Alignment editor (v 7.0.5.3) and Geneious Pro (5.3.4-6.1.5). For each plasmid the forward and reverse reads were assembled into a single contiguous sequences (contig) using the CAP3 sequence assembly program (Huang and Madan, 1999) preinstalled on BioEdit and Geneious Pro. Multiple sequence alignments were done using the ClustalW sequence alignment program (Larkin et al., 2007) in BioEdit or using the Multiple Sequence Alignment (MUSCLE) program (Edgar, 2004a, Edgar, 2004b) in Geneious Pro. The *rif* sequences were classified into A and B groups based on the presence of a distinguishing insert in the N-terminal semi-conserved region of *rifA* genes (Figure 1.4) (Joannin et al., 2008), and assigned a sequence label made up of a number followed by A or B depending on which *rif* group they fell into. A sequence identity matrix was computed in BioEdit to identify similar sequences. Sequences that had a score of 99% and above were considered identical and given the same sequence label. For each sample, proportions of unique sequences were computed and plotted on pie chart.

2.9.2. 454 sequence analysis

Analysis of 454 sequence reads was done in collaboration with Thomas Otto and Martin Hunt of WTSI and Etienne De Villiers of KWTRP. The raw reads were extracted from the 454 SFF (Standard flowgram format) files. A Perl script was used to group the reads by their barcode (MID). Two approaches were used to analyse the 454 reads: (1) mapping reads to homologous genomes sequenced using Illumina done by T. Otto (Appendix method 9.1) and (2) *de novo* assembly of 454 reads into contigs using done by E. de

Villiers (Appendix Script 6.1: Parameters for Newbler assembly of 454 amplicon sequence reads into contigs).

2.9.2.1. Analysis by mapping reads to homologous genomes

For a subset of isolates with whole genome sequence data (Appendix Table 9.6), reads of each bar code were mapped against their reference genome, using SMALT (ftp://ftp.sanger.ac.uk/pub/resources/software/smalt/) with the following parameters: index -k 13 -s 3 (k-mer length of 13 and a step size of 3), mapping -r 0 (random placing of reads) -y 0.5 (identity of complete read must be at least 50%). The resulting output (sam file) was transformed with samtools (Li et al., 2009) to a bam file. In Artemis (Rutherford et al., 2000) BAM view (Carver et al., 2013), RPKM (reads per kilo base pair per million) and read counts were calculated for all the annotated *rif* and *stevor* genes. Read counts were used for analysis of expression because the amplicon lengths were similar hence there was no need to normalize for gene lengths by using RPKM. Read counts were tabulated in Excel and pie charts generated based on the proportion of read counts for each gene across all time points for each isolate. Genes with less than 5X coverage (less than an average of 5 reads across time points) were removed from the list before further analysis in R Studio. Read count data normalization for library size and differential expression analysis was carried out using DESeq (Anders and Huber, 2010) in R Studio. Heat maps were generated in R Studio.

2.9.2.2. De novo contig assembly

For the subset of Kilifi isolates that had no whole genome sequence data, 454 reads were assembled into contigs using the Newbler assembler (Margulies et al., 2005). Newbler was run with default parameters, minimum input read length of 20 base pairs (bp),

minimum overlap length of 40 bp and minimum overlap identity of 90%. The contig length cutoff was set at 100 bp (Appendix Script 6.1: Parameters for Newbler assembly of 454 amplicon sequence reads into contigs).

The contigs were reoriented to all be in the same direction (5'-3') and primer sequences tagging the ends of each amplicon were removed. Non-target amplicons were removed by running MegaBLAST against all *rif* / all *stevor* databases on Geneious Pro. These databases were compiled in house and consisted of 2450 rif and 735 *stevor* sequences from a collection of laboratory (3D7, HB3, Dd2, IT) and wild isolate genomes (from Brazil (Albrecht et al., 2006), Gabon (Abdel-Latif et al., 2002), and Kenya ((Blythe et al., 2009), WTSI, Otto et al, unpublished, www.sanger.ac.uk). The results were binned into hit or no hit. The hit bins were further evaluated using alignments and trees, to remove any non-target sequences.

The cleaned *rif* and *stevor* target contig sets were then clustered to obtain a list of unique *rif* and *stevor* variants from each isolate. A combination of CD-HIT (Huang et al., 2010, Li and Godzik, 2006) and BLAST were used. *Rif* and *stevor* sequences were clustered at 97% and 95.3% identity respectively. The number of reads per contig was computed to give the expression value. Pie charts were generated from the reads data in Excel. DESeq analysis and heat maps were done in R Studio.

2.10. Northern blotting

Northern blotting of RNA from laboratory and field isolates was carried out in order to confirm whether *rif* messenger RNA was intact and full-length, and to estimate expression levels. A non-radioactive method based on digoxygenin (DIG) labeling (Roche) was employed. 1-2% agarose gels containing 2% formaldehyde (Sigma Aldrich)

were prepared by dissolving the appropriate amount of agarose in 141.9ml 1 x MOPS (200mM MOPS, 50mM NaAC, 20mM EDTA) and boiling the mixture. The agarose solution was let to cool to about 55^0C before 8.1ml of 37% formaldehyde was added and the gel let to set. 5µg total RNA was resolved on formaldehyde gels, transferred to positively charged membranes by capillary transfer and fixed by UV crosslinking. Previously published primers were used to generate the probes; *rif*, *varC* (exon2), *msp1*, *calmodulin* and *var* (exon1 DBLα) fragments (Kyes et al., 2000) for DIG-labelling, from genomic DNA and cDNA. DIG-labeled sense and anti-sense RNA probes were generated by *in vitro* transcription using T7 RNA polymerase using the DIG RNA Labeling Kit (SP6/T7) (Roche). *var* and *msp1* were included to check the stage of the parasite while *calmodulin* which is a housekeeping gene was used to check mRNA level across the IDC. The probes were hybridized to the membrane overnight after which high stringency washes were performed. DIG-labeled hybrids were detected with anti-DIG-alkaline phosphatase conjugate and the substrates CSPD or CPD (Roche). Visualization was done under white light on a gel imaging system (Biorad) at increasing exposures to a maximum exposure of 1 hour.

2.11. Sequence diversity analysis

2.11.1. Collection of *rif* and *stevor* sequences from sequence databases

We have obtained the full repertoire of *rif* and *stevor* sequences from 3D7 and IT genome (www.plasmodb.org; www.genedb.org), HB3 genome from the Wellcome Trust Sanger Institute (unpublished), partial *rif* and *stevor* repertoires from the Dd2, HB3 genome sequences (www.broad.mit.edu, Broad Institute of Harvard and MIT), as well as several

rif sequences from a limited number of Brazilian (Albrecht et al., 2006), Gabonese (Abdel-Latif et al., 2002) and Kenyan isolates (Blythe et al., 2009) (Table 3.1).

2.11.2. Multiple sequence alignments

Multiple sequence alignments were carried out using MUSCLE (Edgar, 2004b) or ClustalW in Geneious, or directly on the European bioinformatics web server (http://www.ebi.ac.uk/Tools).

2.11.3. Phylogenetic tree construction and visualization

Maximum likelihood phylogenetic trees constructed from MUSCLE-aligned sequences using PHYML (http://www.atgc-montpellier.fr/phyml) (Guindon et al., 2005, Guindon and Gascuel, 2003). PhyML was run with either of the two substitution models, Generalized Time Reversible (GTR) for nucleotide sequences and LG model for amino acid sequences. Bootstrap analysis was performed with 100 to 500 replicates.

2.12. Live immunofluorescence assays (IFA) to show RIFIN surface localization

2.12.1. Protein expression and antibody purification

The hypervariable region (HVR) of the strain transcendent RIFIN PFL2585c (PF3D7_1253700) was expressed as a FLAG-tagged protein (Kibwana E., Kadasia K. et al., manuscript in preparation) using the cell free wheat germ system (Takeo et al., 2009). The purified antigen was used to identify highly reactive sera from immune adults using enzyme-linked immunosorbent assay (ELISA). The identified highly reactive serum was used for anti-RIFIN HVR antibody purification. The recombinant RIFIN HVR was coupled to Aminolink® beads and incubated with the highly reactive serum overnight to bind anti-RIFIN IgG. The antibodies were eluted from the antigen-beads complexes using IgG elution buffer.

2.12.2. Immunofluorescence assay

Parasites for use in the IFA assay were first assessed for PFL2585c gene expression by RT-PCR. Parasites that were positive for expression were cultured as described above (section 2.3). Cultures were synchronized and grown to above 5% parasitemia before IFA was performed. Early to mid trophozoite parasite stages were targeted for IFA. 500µl of cultures were spun down and the pellets washed three times in incomplete media. 5µl of iRBC pellet was mixed with 5µl of the purified human anti-RIFIN HVR antibody (or total immune serum as positive control) and 5µl 1% BSA solution. The mixture was incubated at 37^0C for 30 minutes to an hour. The pellets were then washed three times in 1000ml cold PBS with 1% BSA. The pellets were incubated with secondary goat anti-human IgG Alexa 488 F(ab)' antibody conjugate at 37^0C for 30 minutes in the dark.

During the last 5 to 10 minutes of incubation with secondary antibody the nuclei stain Hoechst at a final concentration 0.5µg/ml was added. The pellets were then washed three times in 1000µl of cold 1% BSA in PBS and resuspended in 50µl of 1% BSA in PBS. 20µl of the pellets were placed on a glass slide and sealed with a coverslip. The slides were examined using 40x or 100x (oil) objective, and incident light of 450-495nm for FITC fluorescence, 390-440nm for DAPI /Hoechst.

CHAPTER 3

3. Characterising the sequence diversity of *rif* and *stevor* multigene families.

3.1. Introduction.

Plasmodium species contain several large multigene families located in the subtelomeric regions that encode variant surface antigens. One large multigene family called *Plasmodium* interspersed repeat (*pir*) gene family is found in both human and simian malaria parasites as well as rodent malaria parasites (del Portillo et al., 2001, Fischer et al., 2003, del Portillo et al., 2004, Janssen et al., 2004, Pain et al., 2008, Otto et al., 2014a, Otto et al., 2014b). These genes code for proteins that are potentially expressed on the surface of infected erythrocytes and may play important roles in immune evasion of the IE and pathogenesis (Cunningham et al., 2010, Jemmely et al., 2010, Bernabeu et al., 2012, Niang et al., 2014, Niang et al., 2009, Goel et al., 2015). In *P. falciparum,* the *pir* multigene family is made up of two related superfamilies *rif* (Fernandez et al., 1999, Kyes et al., 1999, Cheng et al., 1998) and *stevor* (Limpaiboon et al., 1990, Cheng et al., 1998). The pir genes, together with the well-known var genes (Su et al., 1995, Baruch et al., 1995, Smith et al., 1995, Gardner et al., 2002), encode the major variant surface antigen (VSA) that modify the antigenic and immunogenic properties of infected erythrocytes (IE).

The *var* encoded PfEMP1 is the most well studied VSA. The *var* gene family is highly diverse partly due to recombination and this diversity complicated approaches to classify these genes into meaningful groups. The completion of the first *P. falciparum* genome revealed genetic structuring of *var* genes (Gardner et al., 2002). Based on the 5' flanking

region (upstream) and chromosomal location *var* genes could be classified into 3 major groups, upsA (group A), upsB (group B), upsC (group C), and a minor group upsE *var1-3* (Gardner et al., 2002, Lavstsen et al., 2003). These groups are functionally important (Robinson et al., 2003, Kraemer and Smith, 2003). The semiconserved DBL alpha (DBLα) domain provides a region that can be used to design universal primers to amplify all *var* genes, the DBLα tag (Taylor et al., 2000). Approaches to classify *var* genes based on the DBLα tag region have been developed, (1) based on number of cysteine residues and the presence or absence of mutually exclusive motifs, REY and MFK (Bull et al., 2005) and (2) based on recombining blocks (Bull et al., 2008). The DBLα tag classification is able to provide clinically important information about the full length PfEMP1 molecule. Expression of groupA like *var* tags has been associated with the severe malaria (Warimwe et al., 2009). Because the DBLα tag only covers a very small region of the *var* gene, a classification approach based on full length var genes was recently developed (Rask et al., 2010). In this approach, DBL and CIDR domains were classified using homology blocks, and the sequence in which the domains occurred enabled subgrouping of *var* into domain cassette (DC) groups (Rask et al., 2010). Subsequent studies have associated expression of certain DC types with binding phenotypes and severe malaria (Avril et al., 2013, Turner et al., 2013, Claessens et al., 2012). The many approaches developed to classify *var* genes are a reflection of the difficulties with classification of highly variable multigene families, such as the *var* and *pir* multigene families.

The *pir* multigene family encodes VSA that are coexpressed on the surface of IE together with PfEMP1. The *pir* genes are related in that they share similar gene architecture composed of two transmembrane domains flanking a hypervariable region. This is suggestive of shared ancestry. A previous study suggested that *pir* genes from human and rodent malaria parasites share distant but common ancestry based on sequence homology within the intron as well as shared predicted secondary structure of the proteins they encode (Janssen et al., 2004, Jemmely et al., 2010). *P. falciparum pirs, rif* and *stevor,* share similar architecture and share some sequence homology at the 5' end of the coding sequence and in the 3' region encoding the cytoplasmic tail (Gardner et al., 1998).

The highly diverse *rif* family forms the largest multigene family with 184 copies in the 3D7 genome (http://www.genedb.org/genedb/malaria/). The *rif* family is made up of two major groups *rifA* and *rifB* based on a 25 amino acid insertion in *rifA* (Joannin et al., 2008, Petter et al., 2007, Joannin et al., 2011). This subgrouping was shown to be conserved across laboratory isolates and a few (n=9) wild isolates from Brazil (Albrecht et al., 2006) and Gabon (Abdel-Latif et al., 2002). The *rifA* subgroup is more diverse than *rifB* and contains a majority of the sequences (129/184 in 3D7). The conserved *rifA* that are oriented head-to-head with a neighbouring Group A *var* (Lavstsen et al., 2003) have been named *rifA1* and those oriented head-to-head with UpsA-linked Type 3 var genes are named *rifA3* (Wang et al., 2009). The Type 3 var genes are also highly conserved across parasite isolates (Kraemer and Smith, current opinion micro 2006). Another conserved *rifA* group is the *rifA2* which occurs as a single copy gene in all *P. falciparum* genomes (Wang et al., 2009).The relatively conserved *rifB* subgroup can be further sub-

divided into at least 3 groups, B1, B2 and B3, with the B3 group being more diverse (Joannin et al., 2008, Joannin et al., 2011).

The more conserved *stevor* multigene family has 40 gene copies in the 3D7 genome (Gardner et al., 2002, http://www.genedb.org/genedb/malaria/). Phylogenetic analyses have classified *stevor* into two main groups, a major group, which contains a majority of the sequences and a more diverse minor group (Blythe et al., 2009, Schreiber et al., 2008, Joannin et al., 2011). An even more divergent group within the minor group was observed in a stevor phylogenetic tree derived from multiple alignments of stevor from 3D7, DD2 and IT (Joannin et al., 2011). This group of *stevor* sequences that did not cluster with the rest of the sequences was termed *stevor*-like. This grouping is observed across laboratory isolates and a few isolates from Kilifi on the Kenyan coast (Blythe et al., 2009). The 5' untranslated regions (UTRs) have not been characterized.

With the advent of high throughput sequencing technology more laboratory and wild isolates have subsequently been sequenced. In addition previously sequenced genomes have been resequenced and reannotated to correct gene models. Phylogenetic analysis with the additional cleaner sequence data is necessary in order to validate previous *rif* and *stevor* subgrouping and to determine whether these groups are conserved in wild isolates.

3.2. Objectives

i.) Utilize phylogenetic trees to cluster *pir* genes from *P. falciparum* parasites from different geographical locations into groups.

ii.) Identify conserved *pir* groups in *P. falciparum* isolates.

iii.) Conduct phylogenetic analysis of *pir* upstream non-coding regions.

iv.) Analyze length distribution of *rif* and *stevor* genes and determine whether genes cluster by length.

3.3. Methods.

Some of the methods used in this chapter have been described in the Methods chapter (2.11).

3.3.1. Retrieval of *pir* repertoires from laboratory and wild isolates.

Publicly available *rif* and *stevor* sequence were retrieved from sequence databases (Table 3.1); 3D7 and IT *pir* sequences was obtained from GeneDB (http://www.genedb.org/genedb/malaria/). Dd2 and a partial list of HB3 *pir* sequences were downloaded from the Broad Institute (http://www.broad.mit.edu). With the completion of the HB3 genome sequencing by the Sanger Institute more complete and updated HB3 *pir* sequences were obtained from the Sanger HB3 genome (http://www.sanger.ac.uk/Projects/P_falciparum). Incomplete *pir* repertoires from wild isolates were obtained from published data: *pir* from Brazilian isolates (Albrecht et al., 2006), *rif* from a Gabon isolate (Abdel-Latif et al., 2002), and *stevor* from Kenyan isolates as well as a few sequences from laboratory isolates (Blythe et al., 2009). Later in the study, genomes of 12 Kilifi (Kenya) patient isolates were sequenced by the Sanger Institute. Partial *pir* repertoires were extracted from these genomes. The data set from the 12 Kilifi genomes was only used for identification of *pir* orthologs in wild isolates, and not included in the phylogenetic analysis.

The non-coding upstream regions of 3D7 and IT parasite isolates were obtained from GeneDB by downloading 1kb upstream from the start codons of *pir* genes.

Table 3.1. *Complete and incomplete pir repertoires from different geographical locations.*

Parasite isolate	Geographical location	Whole genome sequence	Complete or partial *pir* full length and amplicon sequences	
			No. of *rif*	No. of *stevor*
3D7[1]	Africa (West Africa)	Y	184 (complete)	42 (complete)
IT[1]	South America (Brazil) / Africa	Y	138	43
HB3[2]	America (Honduras)	Y	142	35
HB3[3]	America (Honduras)	Y	103	31
Dd2[3]	SE Asia (Laos/Thailand)	Y	114	31
Brazilian isolates[4]	South America (Brazil)	N	106a	75a
Gabonesse Isolates[5]	Africa (Gabon)	N	13a	n
Kilifi isolates[6]	Africa (Kenya)	N	n	57a
6816[7]	Africa (Kenya)	Y	177	45
8383[7]	Africa (Kenya)	Y	17*	3*
9106[7]	Africa (Kenya)	Y	155	42
9215[7]	Africa (Kenya)	Y	168	39
9605[7]	Africa (Kenya)	Y	177	41
9626[7]	Africa (Kenya)	Y	19*	2*
9775[7]	Africa (Kenya)	Y	35*	9*
10668[7]	Africa (Kenya)	Y	179	40
10936[7]	Africa (Kenya)	Y	181	41
10975[7]	Africa (Kenya)	Y	27*	9*
11014[7]	Africa (Kenya)	Y	17*	2*
11019[7]	Africa (Kenya)	Y	197	48
SA075[7]	Africa (Kenya)	Y	95#*	31#*

1: PlasmoDB (http://www.plasmodb.org/plasmo/home.jsp)

2: HB3 sequenced at the Sanger Institute (http://www.sanger.ac.uk/Projects/P_falciparum) (Unpublished)

3: Dd2 and HB3 sequenced at the Broad Institute (http://www.broad.mit.edu)

4: pir amplicon sequence data (Albrecht et al., 2006)

5: stevor amplicon sequence data (Abdel-Latif et al., 2002)

6: stevor amplicon sequence data (Blythe et al., 2009)

7: Kenyan patient isolates (12 from Kilifi and 1, SA075, from Kisumu) sequenced at the Sanger Institute (http://www.sanger.ac.uk/Projects/P_falciparum) (Unpublished)

** - Noticeably incomplete pir repertoire due to insufficient sequence data leading to partial whole genome assembly*

 # - Additional 55 rifs and 9 stevors annotated manually

a - amplicon sequences

n – none

3.3.2. Phylogenetic sequence analysis

Multiple sequence alignments were carried out using the program Multiple Sequence Comparison by Log-Expectation (MUSCLE) (Edgar, 2004a) both online on the European Bioinformatics Institute server (http://www.ebi.ac.uk/Tools/msa/muscle/) and within Geneious® (Biomatters, Inc) software (http://www.geneious.com, Kearse et al., 2012), using default settings. Alignments were visualized and edited in Geneious. Phylogenetic analyses were carried out either on the online ATGC bioinformatics platform (South of France bioinformatics platforms) (http://www.atgc-montpellier.fr) or in house within the Kilifi Bioinformatics server. Maximum likelihood phylogenetic trees were constructed from aligned sequences using PhyML (Guindon et al., 2009) with the Generalized Time Reversible (GTR) substitution model for nucleotide sequences and LG model for amino acid sequences. For branch support statistics the aLRT statistic was computed or a Bootstrap analysis performed with 100 to 500 replicates.

3.3.3. Genomic locations of *pir* genes.

Centromere locations, chromosome sizes and chromosomal locations of *pir* genes were obtained for 3D7, IT and the chimpanzee malaria parasite *P. reichenowi* from GeneDB and PlasmoDB. Whole genome sequence data files for the 3 isolates were downloaded from PlasmoDB. The chromosomal annotations of *pir* genes and chromosome lengths were extracted in Geneious as csv files. Centromere locations were obtained from GeneDB. The centromere locations were based on GC-content and have not been experimentally verified. Chromosomal maps were drawn using R studio.

3.4. Results.

3.4.1. Multiple sequence alignments of *rif* and *stevor* genes showing key sequence features.

Multiple alignments were performed on the complete repertoire of *rif* and *stevor* amino acid coding sequences (CDS) from 3D7 for initial characterization of the gene families and to identify key gene features. As has been shown previously the two gene families share similar architecture, a semiconserved N-terminal, two predicted transmembrane (TM) domains flanking a hypervariable region (HVR) that is thought to form a loop on the IE surface, and a conserved intracellular C-terminus (Figures 1.2, 3.1 and 3.2). The *rif* genes have two variable regions, an N-terminal (or 5') variable domain (V1) between the signal sequence and the PEXEL motif and a second hypervariable region (V2/ HVR) between the two predicted transmembrane domains (Figure 3.1) (Joannin et al., 2011). The two subgroups *rifA* and *rifB* were distinguished mainly by the insert within the first conserved region in the N-terminus in *rifA* (Figure 3.1). The *rifA* exhibit higher sequence diversity with 37% overall amino acid pairwise identity as compared to 40% for *rifB* (Table 3.2). Conserved cysteine residues were observed as previously described (Joannin et al., 2008), 10 in *rifA*, and 6 in *rifB* with 5 of them overlapping between *rifA* and *rifB* (Figure 3.1).

Alignments of the more conserved *stevor* gene family showed a slightly different architecture, with one variable region, and the HVR between the two putative transmembrane domains (Figure 1.2 and 3.2). Two residues, a tyrosine and leucine that lie within the conserved region were perfectly conserved in IT and 3D7. One of the

tyrosine residues was conserved in both *rifA* and *rifB*. 7 conserved cysteine residues were observed with 3 being specific to the major *stevorA* group.

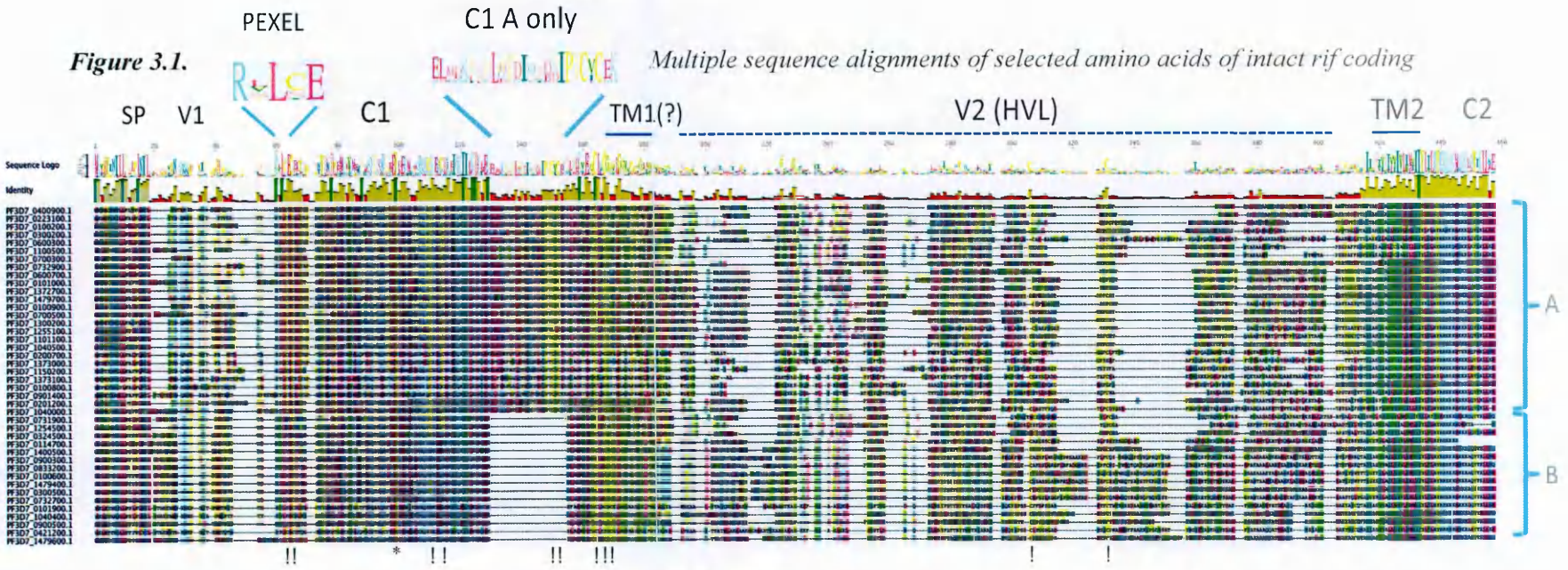


Figure 3.1. *Multiple sequence alignments of selected amino acids of intact rif coding sequences (n=42) from 3D7 highlighting the domain architecture the signal peptide, a highly variable region (V1), PEXEL motif, conserved region, a 25 amino acid insert in C1 of A-RIFIN, a weakly predicted transmembrane domain (TM1), the hypervariable loop/ region (V2/ HVR), a second transmembrane domain (TM2) and a conserved C terminal region (C). The sequence logo is displayed at the top of the alignment. Below the logo is the identity, the mean pairwise identity of all the pairs in the column, Green bars indicate 100% identity, brown-green bars 30-99%, and red less than 30% identity. The exclamation marks (!) indicate positions of the conserved cysteine residues. The asterisks (*) indicate tyrosine residue that is 100% conserved in both rifA and rifB. The major*

rif subgroups rif and rifB are indicated on the right. The percent sequence identities of the different sequence regions are displayed in the table below (Table 3.2).

Table 3.2. *Mean percent pairwise identity of rif amino acid CDS sequences from four genomes, 3D7, IT, HB3 and DD2*

***rif* (575 sequences)**		**Mean % pairwise identities**			
3D7=183, IT=136, HB3=142, DD2=114		**Overall**	**N-terminal**	**HVL**	**C-terminal**
Amino acid sequences	Total	34	45	16	67
	rifA	37	52	16	69
	rifB	40	55	20	70
Nucleotide sequences	Total	46	56	28	72
	rifA	47	60	28	74
	rifB	58	75	37	81

3D7: 183 sequences of 3D7 genome version 3 from GeneDB (http://www.genedb.org/genedb/malaria/)

DD2 -114 sequences from Dd2 genome from the Broad institute (http://www.broad.mit.edu)

HB3: 142 sequences from HB3 genome from WTSI (ftp://ftp.sanger.ac.uk/pub/project/pathogens/Plasmodium/falciparum/HB3/, 2013)

IT: 136 sequences from IT version 3 genome on GeneDB (http://www.genedb.org/genedb/malaria/)

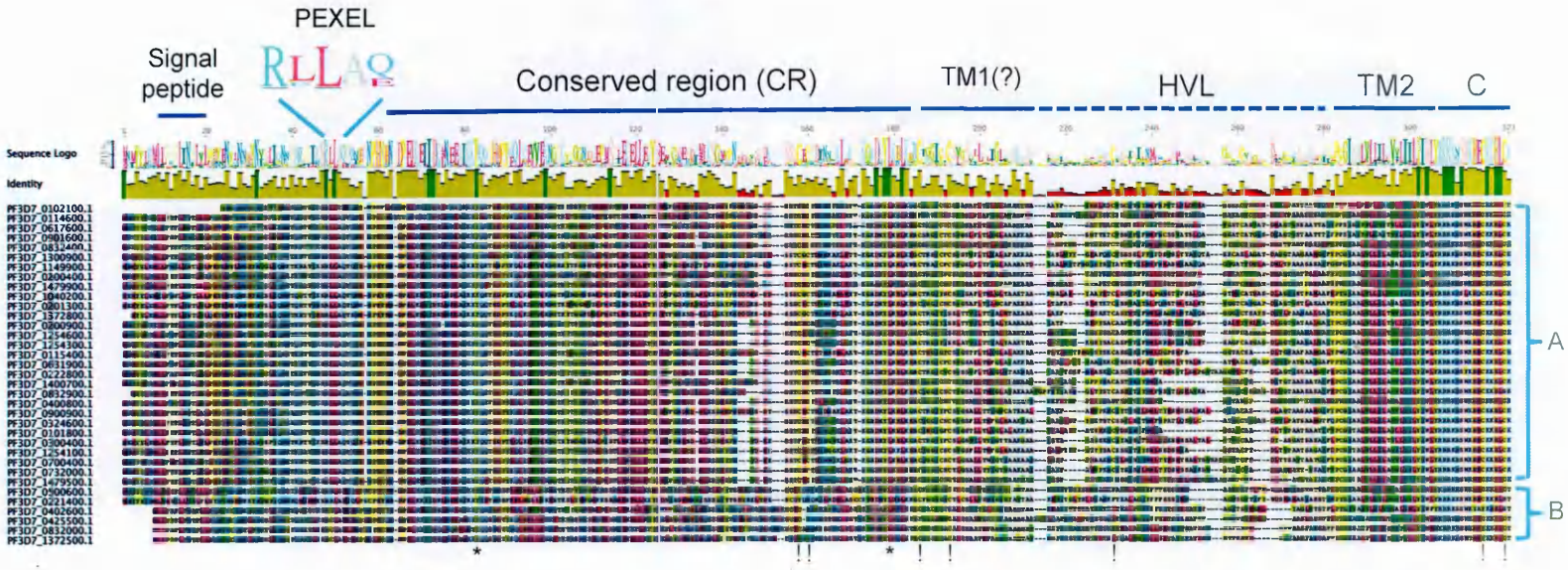


Figure 3.2 *Multiple sequence alignments of the amino acids of intact stevor coding sequences (n=35) from 3D7 highlighting the domain architecture the signal peptide, PEXEL motif, conserved region, two putative transmembrane domains, a first transmembrane domain (TM1) which is not as strongly predicted as the second (TM2), the hypervariable loop/ region (HVR) and a conserved C terminal region (C). The sequence logo is displayed at the top of the alignment. Below the logo is the identity, the mean pairwise identity of all the pairs in the column, Green bars indicate 100% identity, brown-green bars 30-99%, and red less than 30% identity.*

The asterisks (*) indicate tyrosine and leucine that are 100% conserved in 3D7 and IT. The exclamation marks (!) indicate positions of conserved cysteine residues. The major (A) and minor (B) are indicated on the right.

Table 3.3. *Mean percent pairwise identity of stevor amino acid and nucleotide CDS sequences from four genomes, 3D7, IT, DD2, HB3 and IT.*

***stevor* (146 sequences)**	**Mean % pairwise identities**			
3D7=42, IT=40, HB3=35, DD2=29	**Overall**	**N-terminal**	**HVL**	**C-terminal**
Amino acids	47	51	21	72
Nucleotides	61	66	40	75

3D7: 42 sequences of 3D7 genome version 3 from GeneDB (http://www.genedb.org/genedb/malaria/)

DD2 -29 sequences from Dd2 genome from the Broad institute (http://www.broad.mit.edu)

HB3: 35 sequences from HB3 genome from WTSI (ftp://ftp.sanger.ac.uk/pub/project/pathogens/Plasmodium/falciparum/HB3/, 2013)

IT: 40 sequences from IT version 3 genome on GeneDB (http://www.genedb.org/genedb/malaria/)

3.4.2. Characterising *rif* and *stevor* sequences relationships using phylogenetic trees.

To explore *rif* and *stevor* repertoire diversity phylogenetic trees were constructed from available genome sequence data. A maximum likelihood tree from 224 MUSCLE-aligned *rif* and *stevor* coding sequences (CDS) from 3D7 was constructed to show the phylogenetic relationships between members of the two multigene families (Figure 3.3). The tree showed clustering of the two families into separate but related groups. Previously described *rif* and *stevor* major subgroups, A and B were observed and *rifB* subgroups *rifB1*, *rifB2* and *rifB3* (Joannin et al., 2008). The *rifA* group was further clustered into 10 subgroups, A1 to A10. The *rifA1*, *rifA2* and *rifA3* groups have been previously described (Wang et al., 2009), although in our case the A1 and A3 naming is not strictly applied to the UpsA *rif* sequences as previously described. In our case the naming was given to clusters containing these special *rif* genes from the phylogenetic tree. A few other non-UpsA *rif* sequences clustered with the previously defined A1 and A3 UpsA var-linked *rif* genes thus making these groups bigger in our classification (Figure 3.3). The previously described strain-transcendent *rif* (PFD0070c, previously named rifA2 (Wang 2009), together with its orthologs in Dd2 (PFDG_04771) and HB3 (PFHG_03700) and the related 3D7 pseudogene PFL2580w formed a divergent but extremely well-conserved group named A2. A few *rifA* sequences did not fall into any group and one *rifA* clustered with *rifB*. The *rifB* further clustered into three groups B1, B2 and B3 as previously described.

The *stevor* gene family subdivided into two types, the major group with a majority of the sequences (named *stevorA*), and a more divergent minor group (*stevorB*) (Figure 3.3). The *stevorB* subdivided into four divergent groups while *stevorA* formed four groups.

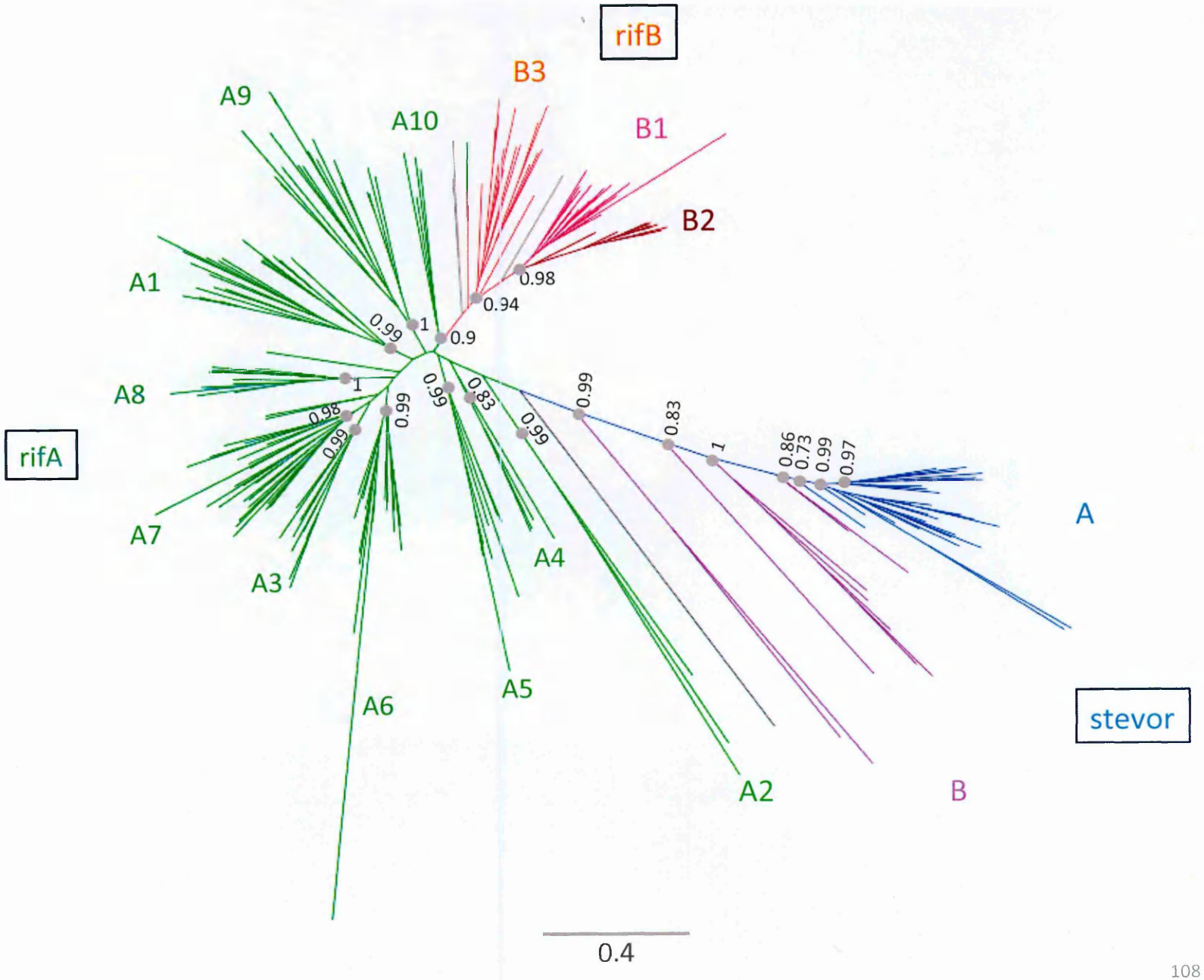

rifB
B3
A9
A10
B1
B2
A1
0.98
0.94
0.99
1
0.9
A8
1
0.99
0.83
0.99
0.83
0.99
0.98
0.99
rifA
1
0.86
0.73
0.99
0.97
A7
A
A4
A3
A6
A5
stevor
A2
B
0.4

Figure 3.3. *Maximum likelihood tree of the 3D7 pir gene family from showing clustering of stevor (blue) and rif genes into separate groups. rif further clustered into two major groups, A (green) and B. Previously described B subgroups (Joannin et al., 2008) are shown, B1 (pink), B2 (brown) and B3 (orange) subgroups (green). rif B that did not fall into either of the B subgroups are shown in black. rifA was further classified into 10 subgroups, with three previously described groups, A1, A2 and A3 (Wang et al., 2009). Previously described stevor A (major) and B (minor) subgroups (Schreiber et al., 2008) are in blue and purple respectively. Branch support indicated is the log likelihood value, the higher the number the more likely the clade. The tree was constructed from MUSCLE-alignments of 224 rif (n=184) and stevor (n=40) coding sequences, constructed using PhyML (run on the ATGC bioinformatics platform South of France: http://www.atgc-montpellier.fr), GTR model, viewed and edited using FigTree v1.4.0. Log-likelihood: - 140966. Scale: nucleotide substitutions per site.*

3.4.2.1. Phylogenetic analysis of *rif* sequences from laboratory and wild isolates.

Generally *rif* and *stevor* repertoires are similar across different isolates. To analyse whether the subgroups were maintained in other laboratory and wild isolates, phylogenetic trees were constructed using *rif* and *stevor* sequences from additional laboratory and wild isolates. The sequences were extracted from genome sequence databases and published literature.

519 *rif* CDS sequences were obtained from publicly available resources, 181 *rif* sequences from 3D7 (PlasmoDB), 113 from Dd2 and 106 from HB3 (Broad Institute of MIT and Harvard), 106 *rif* sequences from nine Brazilian isolate (Albrecht et al., 2006) and 13 sequences from a Gabonese clinical isolate (Abdel-Latif et al., 2002). A phylogenetic analysis of multiple alignments of the 519 *rif* nucleotide sequences resulted in a tree with long branches and very low statistical support (log likelihood of -305,956) because of the high diversity of the recombining *rif* sequences (Figure 3.4). The branch support was low for most of the clades (bootstrap values below 100 out of 500 replicates). Despite the low statistical support an overall tree structure similar to that of the 3D7 only *rif* tree was observed, with *rifA* separating from *rifB* and further subdividing into *rifA* and *rifB* subgroups previously observed (Figure 3.4) with a few cluster shifts. This phylogenetic tree revealed that the major *rif* subgrouping was conserved across genomes (Figure 3.4). The *rifA* grouped into 12 subgroups, with smaller subgroups splitting off previously observed subgroups A2, A5 and A6. The previously described *rifA1* and *rifA3* sequences clustered together with other non-upsA *rif* sequences in this tree (labeled A1 and A3 respectively). The ancient strain-transcendent *rifA2* PFD0070c

and orthologs from DD2 and HB3 were maintained in their own cluster together with the related pseudogene PFL2580w forming an out-group labeled A2a on the tree. This *rif* was not recovered in the field isolates most likely because of lack of coverage by primers. This *rif* is divergent and would not be captured by universal *rif* primers. The *rifB* subdivided into the previously described groups B1, B2 and B3. B3 further subdivided into 3 smaller groups B3a, B3b and B3c (Figure 3.4).

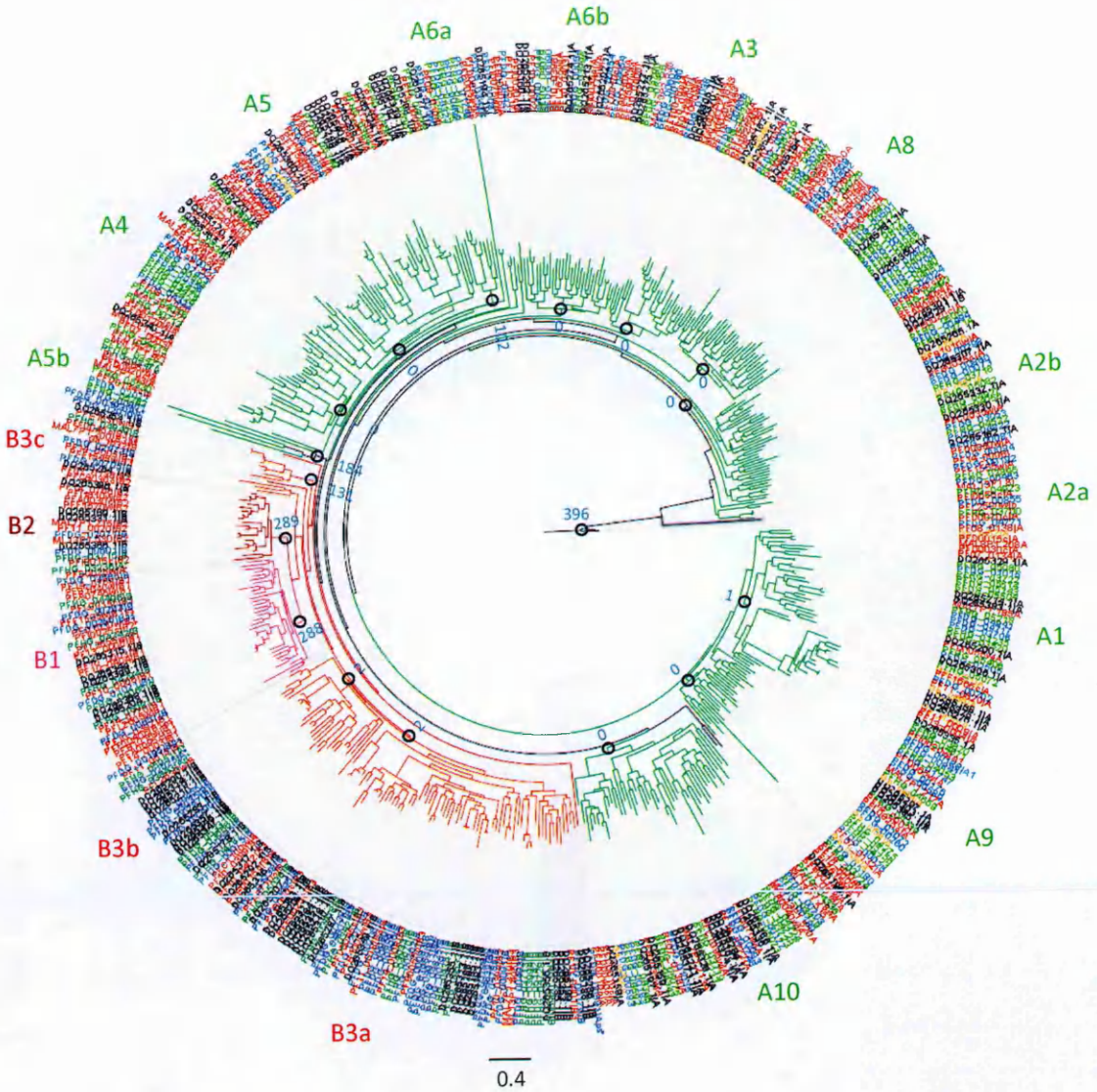

A6a
A6b
A3
A5
A8
A4
A5b
A2b
B3c
A2a
B2
396
289
288
B1
A1
A9
B3b
A10
B3a
0.4

Figure 3.4. *Maximum likelihood tree of rif nucleotide sequences from laboratory isolates 3D7 (red), DD2 (blue), HB3 (green) and partial repertoires from wild two isolates from Gabon (yellow) and Brazil (black), showing clustering of rif from distinct geographical locations into previously described groups (Figure 3.3) with some sequences falling off larger groups and forming smaller clusters. rif A – green, B1 - pink, B2 - brown and B3 – orange branches. The clades are labeled with the bootstrap values (out of 500 replicates). The tree was based on MUSCLE-alignments of 519 rif CDS nucleotide sequences (3D7=181, DD2=113, HB3=106, Brazilian=106 and Gabonese=13). The tree was constructed using PhyML (in house), GTR model, with 500 bootstrap replicates. The tree was viewed and edited using FigTree v1.4.0. Log-likelihood: -305,956.*

3.4.2.2. Phylogenetic analysis of the hypervariable regions of *rif* sequences from laboratory isolates.

The hypervariability of *rif* genes is concentrated within the HVRs, with the N and C terminal regions being relatively conserved (Figure 3.1 and Table 3.2). The HVRs are potentially exposed to host immune responses and are likely to have diversified under immune selection pressure. In addition these regions may be undergoing recombination or gene conversion processes creating sequence mosaics that would further increase the sequence diversity in order to evade the host immune responses. Phylogenetic analysis was carried on *rif* HVRs to determine whether there is any structuring of sequences based on these highly variable region, and to identify putative recombining groups of *rif* genes. For this analysis the HVRs regions were extracted starting from the first transmembrane domain (TM1) up to the second transmembrane domain (TM2), starting and ending with conserved codons encoding cysteine and arginine respectively. Only sequences with long enough sequences running from TM1 up to TM2 were used (3D7=179, DD2=113, HB3=105 and Gabon=13). Sequences from the wild isolates from Brazil were excluded because of missing TM2 sequences.

Pir genes, like any other genes with hypervariable region, present several challenges for clustering using alignment-based traditional phylogenetic approaches. One is the nonuniform rate of evolution, with a higher rate in the hypervariable regions as compared to the conserved regions. In addition, recombination may be occurring in the hypervariable region that is not typically accounted for during phylogenetic

reconstruction. The resulting trees are therefore to be interpreted with caution, and are useful for describing the bigger groups. Nonetheless, there were interesting results obtained from this analysis.

The resultant phylogenetic tree showed clustering of *rif* HVR sequences into 16 groups labeled H1 to H16, 10 *rifA* groups, 2 *rifB* groups B1 and B2, and 4 groups made up of *rifA* and *rifB* sequences (Figure 3.5). The unique ancient *rifA2* PFD0070c with its orthologs from DD2 and HB3 together with the related pseudogene PFL2580w formed a unique cluster in the tree named H1. Most *rifA* HVRs clustered together with only a few grouping with *rifB*. The group H12 contained a majority of *rifA* HVR sequences with others being distributed into the other *rifA* and the mixed *rifA/rifB* groups. H16 contained the *rifA3* HVR sequences together with other non-UpsA *rif* sequences. For *rifB* the two groups B1 and B2 clusters were maintained in the HVR tree (H14 and H13 respectively) while the more diverse B3 group fell into different HVR groups, a majority of the sequences falling into H7 together with a few *rifA*. Other *rifB3* HVRs were observed in H5, a cluster that contained the other UpsA *rifs,* the *rifA1*. H3 and H13 contained the remaining *rifB3* sequences together with *rifA* sequences. The extensive mixing of *rifB3* and *rifA* HVRs was striking and implied that between-group recombination was occurring. In fact in four instances the *rifA* and *rifB* HVRs were in adjacent branches within the same clades. For these four groups a pairwise alignment was done for each pair and pairwise identities were indeed much higher than would be expected for non-orthologous sequences within the HVR, ranging between 71% and 86% (**Table 3.4.1**). Some level of clustering of *rifA* and *rifB* HVRs was not entirely unexpected because the

most apparent difference between the A and B groups is the 75 nucleotide insertion in *rifA* that is upstream of the HVR and was not included in the sequences for the HVR tree. Although there are distinguishable sequence motifs unique to *rifA* or *rifB*, without the most prominent differentiating feature, the insertion in As, it is expected that some *rifA* and *rifB* would group together. The fact that group mixing was only observed for *rifB3* and not *rifB1* and *rifB2* implies that *rifB3* forms a special group that might represent a transition group between A and B types, generated by recombination. The *rifB1* and *rifB2* groups are well conserved in both full length sequences and in the most diverse HVR regions suggesting restricted recombination that occurs only within group. These groups might be evolutionarily constrained to maintain specific functions. The UpsA *var*-linked *rifs* also formed stable groups in both the full-length and HVR tree suggesting limited recombination outside of the group probably to conserve their function.

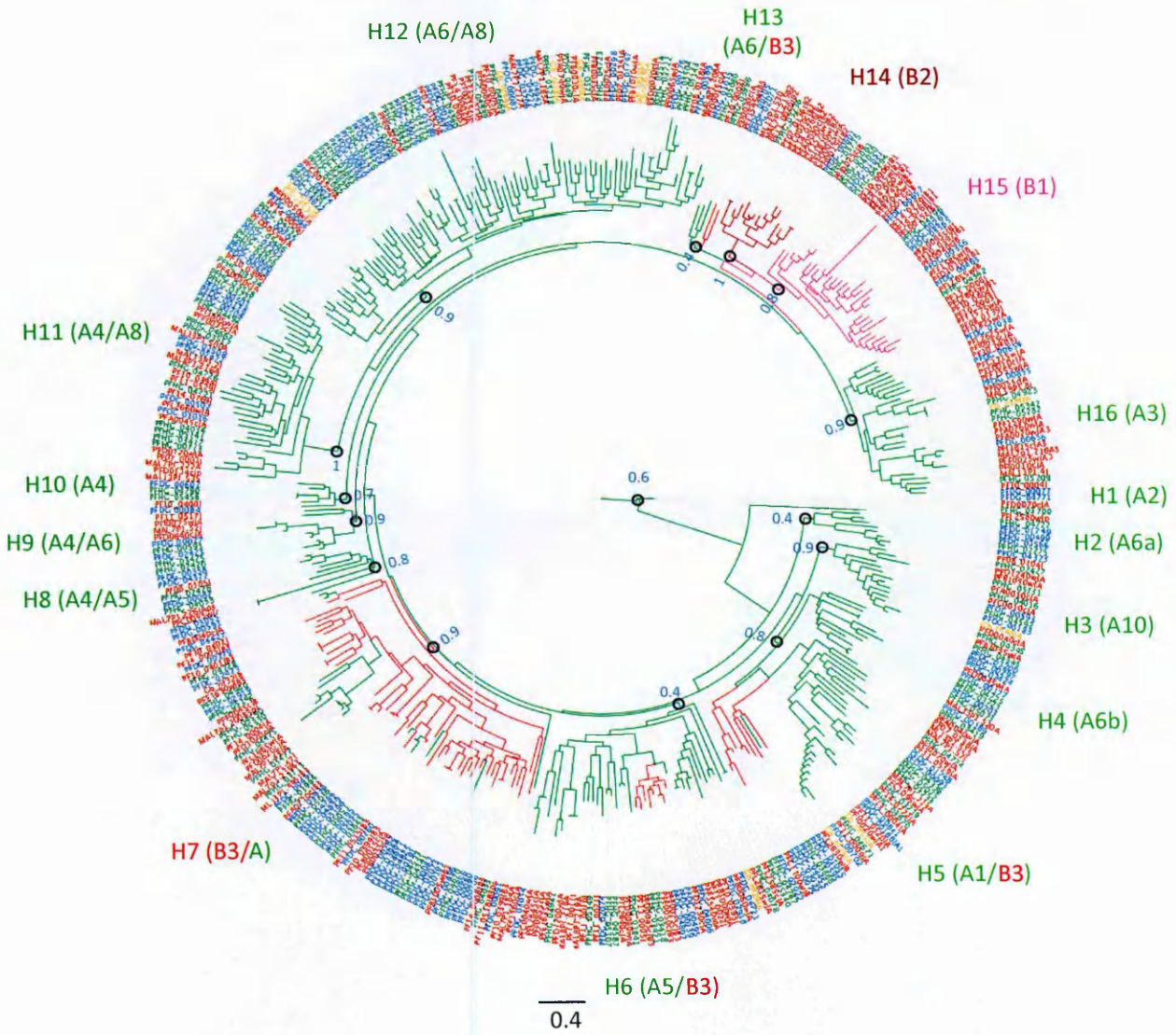

H12 (A6/A8)
H13
(A6/B3)
H14 (B2)
H15 (B1)
H16 (A3)
H1 (A2)
H2 (A6a)
H3 (A10)
H4 (A6b)
H5 (A1/B3)
H6 (A5/B3)
H7 (B3/A)
H8 (A4/A5)
H9 (A4/A6)
H10 (A4)
H11 (A4/A8)
0.4

Figure 3.5*. Phylogenetic tree based on the hypervariable region of rif sequences from 3 laboratory isolates 3D7 (red), Dd2 (blue) and HB3 (green) and a wild isolate from Gabon (yellow). The sequences cluster into clades (H1 to H16). The rifA branches are coloured in green and B in pink (B1), brown (B2) and orange (B3). Previously described clusters from the full-length tree indicated in brackets in the cluster labels, with rifA in green and rifB in red numbers. Branch support indicated is the log likelihood value (blue numbers). The tree was based on MUSCLE-alignments of 411 rif HVR sequences (3D7=181, DD2=112, HB3=105 and Gabonese=13). The tree was constructed using PhyML (run on the ATGC bioinformatics platform South of France: http://www.atgc-montpellier.fr), GTR model, viewed and edited using FigTree v1.4.0. Log-likelihood: -159,763.*

Table 3.4.1. Table showing pairwise percent identities of related rifA and rifB3 HVR sequences clustered in the same clades in the HVR tree.

	ID	Type	% pairwise identity (identical sites over full length)
1	PFDG_00193 PDHG_02073	A" B	81% (445/546)
2	PF10_0394 PFB0015c	B3 A	76% (399/523)
3	PFHG_05051 PFDG_00195	A B*	86% (468/539)
4	PFL2630w PFDG_01035	B A	71% (393/551)

3.4.2.3. Phylogenetic analysis of the semiconserved N terminal regions of *rif* sequences from laboratory isolates.

The phylogenetic tree resulting from full-length *rif* sequences above had long branches and low statistical support because of the hypervariability of the genes. The variability of *rif* and *stevor* genes is concentrated in the HVR with the N and C terminals being semiconserved (Table 3.2 and 3.3). Therefore the trees that result from alignments of full-length sequences have low statistical support and long branch lengths, as observed above (Figure 3.3 and 3.4). To construct phylogenetic trees with better statistical support the N terminal regions of 377 rif genes from 3D7, Dd2, HB3, and a Gabonese isolate were aligned and phylogenetic trees constructed based on this semiconserved region. The resultant tree was structurally different with short branches and laddering of the branches (Figure 3.6). The sequences clustered into 20 groups labeled N1 to N20. The *rifA* and *rifB* groups were resolved into separate clades with *rifA* clustering into 14 groups and *rifB* subgrouping into 6 groups. The *rifA* subgroups were restructured, with sequences from the subgroups identified in the full-length tree shifting into new groups. The *rifA2* divergent group was maintained as an out-group, N20. The *rifA1* and *rifA3*, the UpsA *rif*, clustered into one group, N15, together with a few other non-*upsArifA* genes. The B1, B2 and B3 subgroups were observed. B3 genes were distributed in 5 different subgroups, N10, N11, N12 unique to B3, and N6 and N14 containing *rifA* and *rifB3* sequences. B1 further subdividing into B1a and B1b (N7 and N8). Two small subgroups were made up of *rifA* and *rifB* N terminal sequences (N6 and N14), but besides these there was no mixing of *rifA* and *rifB*. The mixing of *rifA* and *rifB* N terminal sequences was unexpected because one definitive feature that differentiates the two groups is the

presence of a 75 nucleotide insert within the *rifA* N terminal region upstream of the first transmembrane domain, as discussed earlier (Figure 3.1). The observation of closely related *rifA* and *rifB* sequences suggest that some recombination or gene conversion between groups does occur, but at low frequency because only two such groups were observed.

From the analysis it was clear that the N terminal regions formed a different tree structure from the full-length sequences, with the subgrouping becoming distorted as a result of the missing information from the HVR and C terminal regions. Sequence sharing was also observed for *rifA* and *rifB3* in the N terminal regions, despite *rifB* missing the 75 nucleotide insert. As expected the statistical support for the tree was higher for the N terminal tree (*Log-likelihood – 34501)* than for the full-length tree but the overall tree structure was significantly different. These results give support for using full-length sequences for *rif* sequence classification.

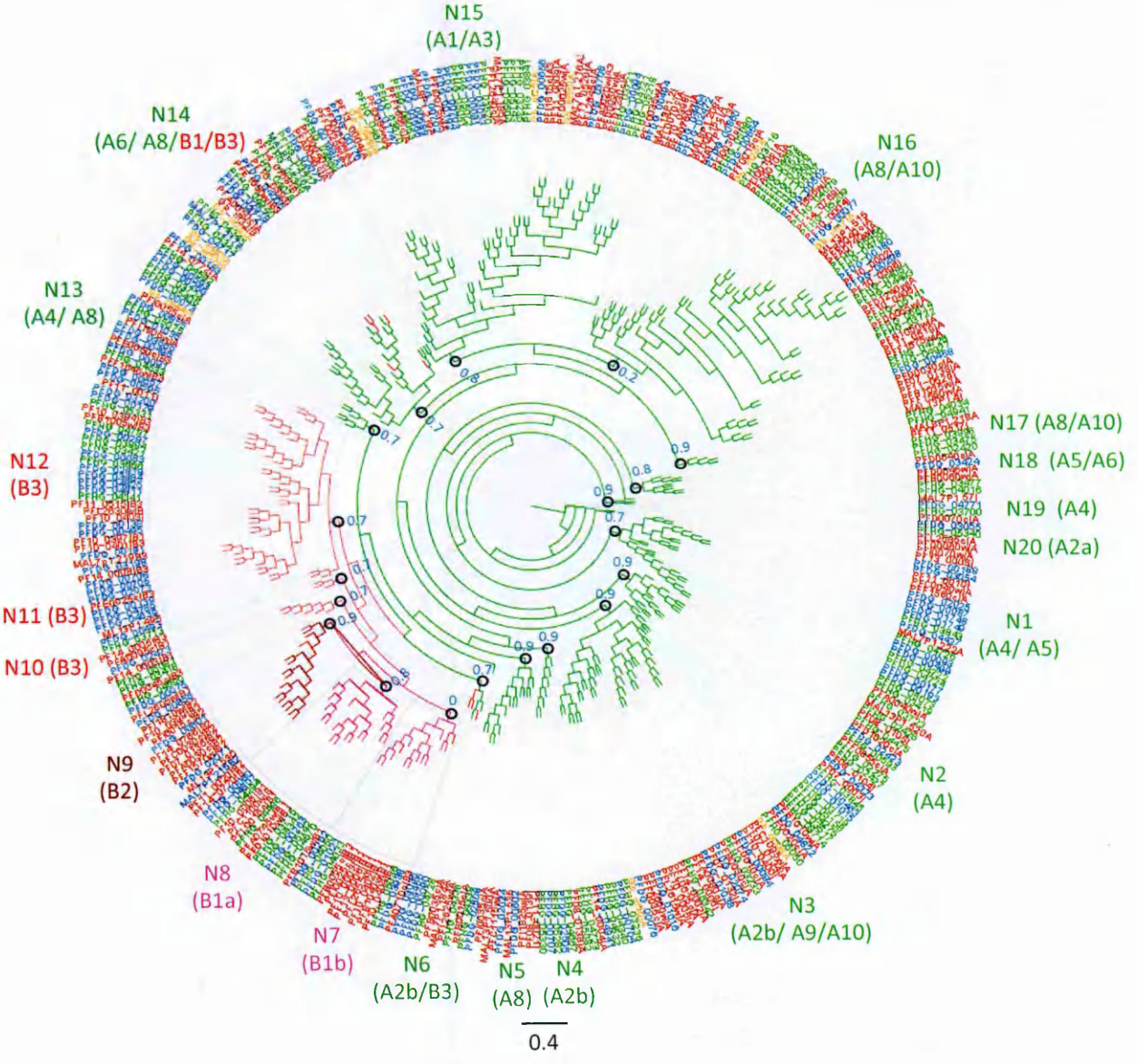

N15
(A1/A3)
N14
(A6/ A8/B1/B3)
N16
(A8/A10)
N13
(A4/ A8)
N17 (A8/A10)
N18 (A5/A6)
N12
(B3)
N19 (A4)
N20 (A2a)
N11 (B3)
N1
(A4/ A5)
N10 (B3)
N9
(B2)
N2
(A4)
N8
(B1a)
N3
(A2b/ A9/A10)
N7
(B1b)
N6
(A2b/B3)
N5
(A8)
N4
(A2b)
0.4

Figure 3.6. *Phylogenetic tree based on the semiconserved N terminal region of rif sequences from 3 laboratory isolates 3D7 (red), DD2 (blue) and HB3 (green) and a wild isolate from Gabon (yellow). The sequences cluster into clades within the A and B groups and are marked with a grey circle, rifA green branches, rifB1 pink, rifB2 brown and rifB3 orange branches. Previously described clusters from the full-length tree (Figure 3.4) are indicated in brackets in the cluster labels, with rifA in green and rifB in red numbers. Branch support indicated is the log likelihood value (blue numbers). The tree was based on MUSCLE-alignments of 377 rif N terminal amino acid sequences (3D7=158, DD2=103, HB3=103 and Gabonese=13). The tree was constructed using PhyML (run on the ATGC bioinformatics platform South of France: http://www.atgc-montpellier.fr), GTR model, viewed and edited using FigTree v1.4.0. Log-likelihood: -34501.*

3.4.2.4. Phylogenetic analysis of *stevor* sequences from laboratory and wild isolates.

To explore phylogenetic relationships of the *stevor* multigene family 102 stevor sequences from laboratory isolates were retrieved from genome databases: 40 sequences from 3D7 (PlasmoDB), 31 sequences from Dd2 and 31 sequences from HB3 (Broad Institute). The sequences clustered into the two main groups (Figure 3.7), the major group *stevorA* and a minor group *stevorB*, as previously observed (Abdel-Latif et al., 2002, Joannin et al., 2011) and in the 3D7 *rif/stevor* tree above (Figure 3.3). The *stevorA* major group was further classified into 13 subgroups A1 to A13 (Figure 3.7). The *stevorA* clusters were characterized by short branches, which is indicative of the sequence conservation of this group. With exception of one group, *stevorA1*, all *stevorA* clusters contained sequences from all the laboratory isolates. A1 did not have any sequences from 3D7. The *stevorB* group was further subdivided into 5 subgroups, B1 to B5 (Figure 3.7). All clusters contained stevor sequences from the different isolates with no isolate-specific cluster, suggesting that stevor repertoire was conserved across isolates from different geographical locations. The previously described *stevor-like* divergent cluster (Joannin et al., 2011) was observed with the members from the three genomes (Cluster B1 in Figure 3.7).

The branch support, log likelihood ratio, for most of the clades was lower than 0.9 but above 0.6.

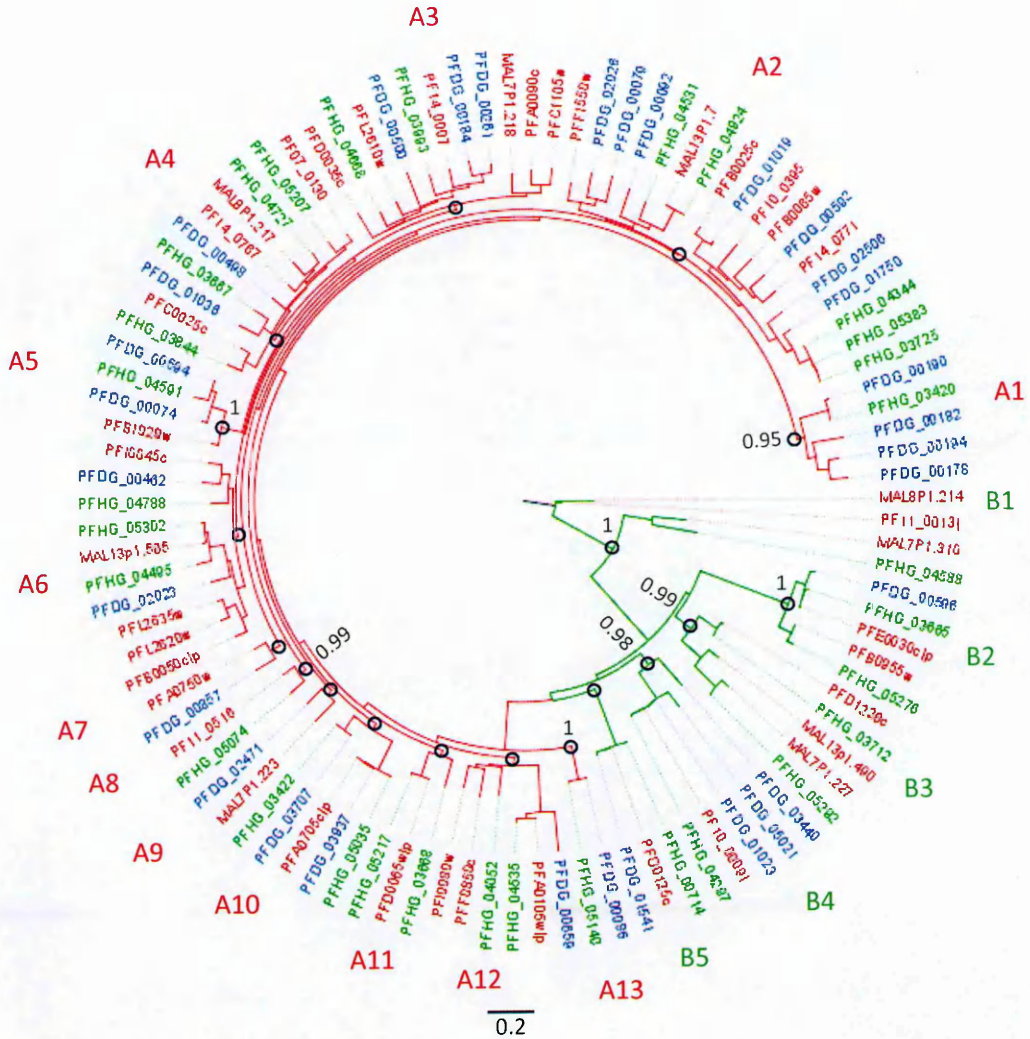

A1
A2
A3
A4
A5
A6
A7
A8
A9
A10
A11
A12
A13
B1
B2
B3
B4
B5
0.95
1
0.99
0.98
0.2

Figure 3.7. *Maximum likelihood tree of 102 stevor nucleotide sequences from laboratory isolates 3D7 (red), DD2 (blue) and HB3 (green) showing the two larger stevor groups the major group stevorA (red branches) and the minor group stevorB (green branches). The larger groups further split into smaller groups. The clades are marked with a black open circle. The stevorA splits into 13 subgroups (A1-A13) and the stevorB into 5 subgroups (B1-B5). Nodes with high branch support (log likelihood >0.9) were labeled. The tree was constructed from MUSCLE-alignments of 102 stevor CDS nucleotide sequences (3D7=40, DD2=31 and HB3=31), using PhyML (run on the ATGC bioinformatics platform South of France: http://www.atgc-montpellier.fr), GTR model. The tree was viewed and edited using FigTree v1.4.0. Log-likelihood: -33,625.*

Analysis of full length *stevor* coding nucleotide sequences (CDS) from laboratory isolates and amplicon sequences from laboratory and Kilifi isolates from published data (Blythe et al., 2009) showed similar clustering as with the laboratory isolates (Figure 3.7), with the subgroups maintained and only minor shifting of clades on the tree (Figure 3.8). The major groups *stevorA* and *stevorB* as well as their subgroups were observed. The *stevorA* further split into 14 subgroups (A1 to A14). None of the amplicon sequences from Blythe, et al. (Blythe et al., 2009) were from the *stevorB* group. This is a divergent group and unlikely to be captured by universal primers. A new clade was observed that was made up of sequences from a Kilifi isolate and the laboratory isolate T9/96 (clade A14 in Figure 3.8). This clade was not previously observed in 3D7, Dd2 or HB3.

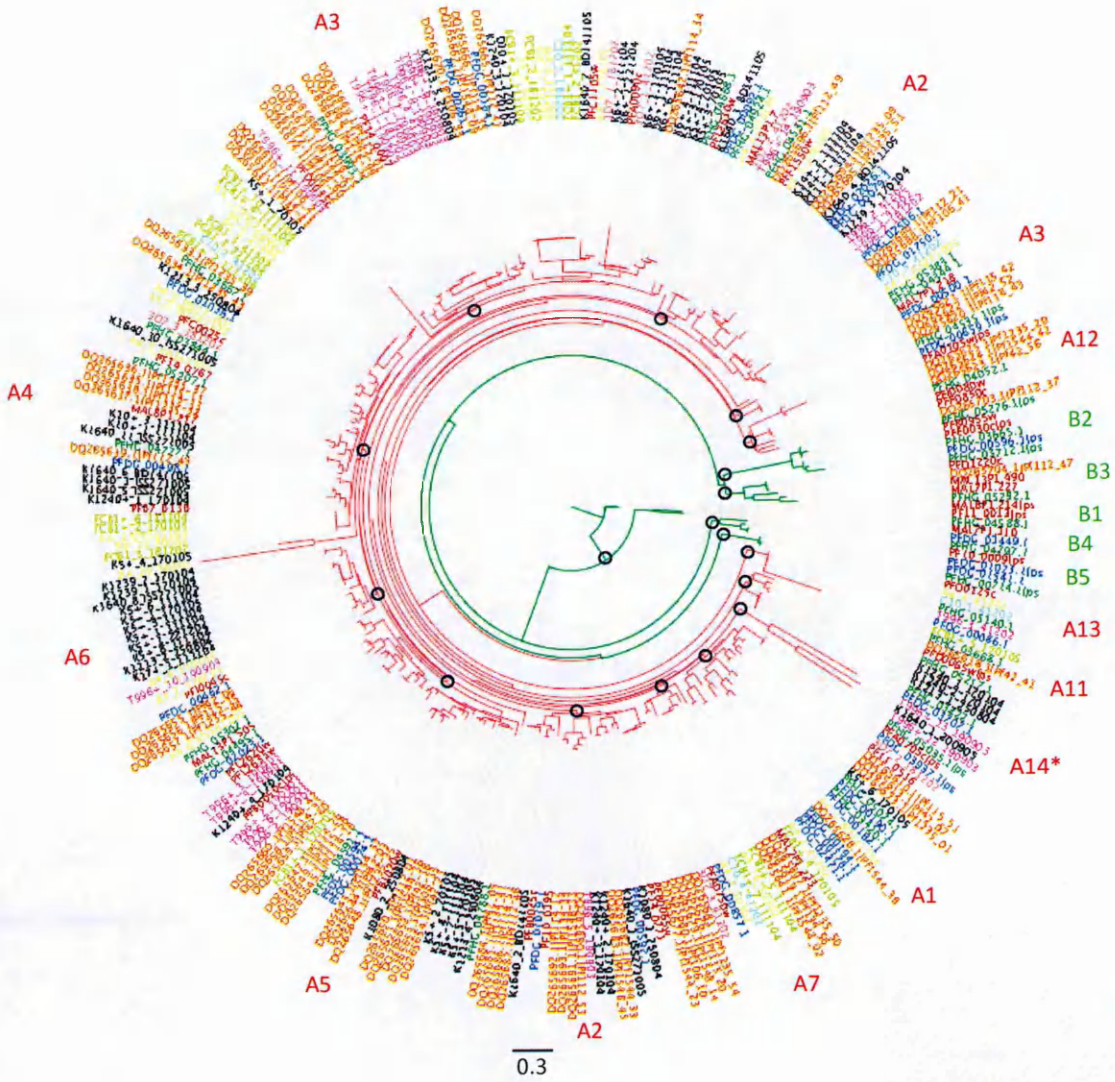

A3
A2
A3
A12
B2
B3
B1
B4
B5
A13
A11
A14*
A1
A7
A2
A5
A6
A4
0.3

Figure 3.8. *Maximum likelihood tree of 308 stevor full length nucleotide sequences from laboratory isolates 3D7 (red), DD2 (blue) and HB3 (green), and a wild isolate from Brazil (orange) and amplicon length sequences from laboratory (light coloured tip labels) and Kilifi isolates (black) from published data (Blythe et al., 2009). The sequences clustered into the two main stevor groups the major group stevorA (red branches) and the minor group stevorB (green branches). The two groups further split into smaller groups (clades are marked with a black open circle), stevorA into 14 subgroups (A1-A14) and stevorB into 5 subgroups (B1-B5). The asterisks (*) indicated a new clade not observed in the tree of stevor sequences from laboratory isolates (Figure 3.7). The tree was constructed from MUSCLE-alignments of 308 stevor CDS nucleotide sequences (full length stevor; 3D7=40, DD2=30, HB3=30, Brazilian isolate=75. Amplicon sequences; 3D7=6, A4=24, C10=5, D10=1, FCB1=17, T9/69= 23 and Kilifi isolates=57), using PhyML (in house), GTR model. The tree was viewed and edited using FigTree v1.4.0. Log-likelihood: -52,415.*

3.4.2.5. Phylogenetic analysis of the hypervariable region of *stevor* sequences from laboratory isolates.

As with the *rif* multigene family *stevor* sequence diversity is concentrated in the HVR. Phylogenetic analysis was carried out based on these regions to identify recombining stevor groups. The HVRs including the two transmembrane domains were extracted from stevor sequences for laboratory isolates (3D7, DD2 and HB3). The sequences were between 117 and 228 nucleotides long. Multiple sequence alignments were done and a phylogenetic tree constructed. The resultant tree was characterized by long branches suggestive of the high variability and poor resulting alignments of the sequences (Figure 3.9). The branch support for most nodes was low (below 80 out of 100 bootstrap values). The stevor HVRs clustered into 12 groups (H1 to H12), 6 *stevorA* only groups and 6 *stevorB* only groups. No A and B mixed groups were observed, indicating lack of between-group recombination *stevorA* and *stevorB* HVR sequence. The *stevorB* HVR clusters were more similar to the full-length clusters with only two exceptions, B1 splits into two H2 and H12, while B3 splits into H1 and H5 groups (Figure 3.9). The *stevorB* groups showed higher bootstrap support than the *stevorA* groups. The *stevorA* HVR showed high within group mixing. All the *stevorA* HVR groups were made up of a combination of sequences from different groups from the full-length tree. This observation suggests that *stevorA* sequences undergo recombination with other members within the group. Overall the grouping observed in the HVR tree was significantly different from the full length tree and this is telling of the different evolutionary pressures working on the different gene regions. The HVR are under diversifying selection pressure hence the high diversity, which may be important for a role in immune evasion of infected erythrocytes.

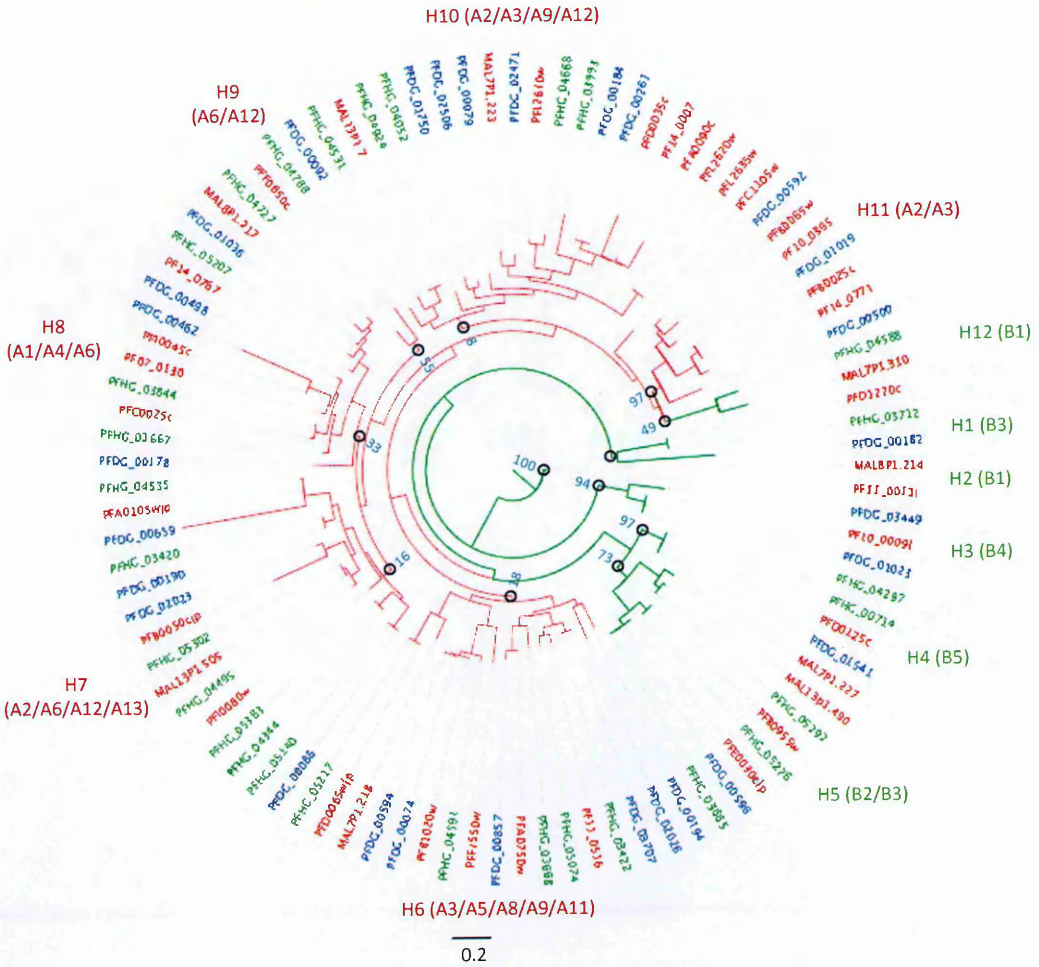

H10 (A2/A3/A9/A12)
H9
(A6/A12)
H11 (A2/A3)
H8
(A1/A4/A6)
H12 (B1)
H1 (B3)
H2 (B1)
H3 (B4)
H7
(A2/A6/A12/A13)
H4 (B5)
H5 (B2/B3)
H6 (A3/A5/A8/A9/A11)
100
94
97
49
97
73
99
99
55
33
16
0.2

Figure 3.9. *Maximum likelihood tree of 97 HVR and the two transmembrane domains of stevor nucleotide sequences from laboratory isolates 3D7 (red), Dd2 (blue) and HB3 (green). The HVR sequences clustered into 12 groups (H1 to H12) with no mixing between the two main groups, stevorA (red branches) stevorB (green branches) separated out. The larger stevorA further split into 6 large clusters (H6 to H11) and the stevorB into 6 smaller groups (H1 to H5 and H12). The clades are marked with a black open circle. Nodes are labeled with the bootstrap value (out of 100 bootstrap replications). The tree was constructed from MUSCLE-alignments of 97 stevor HVR plus TMs nucleotide sequences (3D7=39, DD2=29 and HB3=29), using PhyML (run in house). Substitution model: GTR. Log-likelihood: - 14,468. The tree was viewed and edited using FigTree v1.4.0.*

3.4.2.6. Phylogenetic analysis of the semiconserved region of *stevor* sequences from laboratory isolates.

The N terminal region of stevor genes is highly conserved with overall sequence identity of over 50% for sequences from different parasite isolates (Figure 3.2 and Table 3.3). The full length trees had low likelihood because of the poorly aligned HVRs. Trees based on the more conserved N terminal regions would be less computationally intense and generate trees with better statistical support. This analysis sought to determine whether these regions are phylogenetically informative enough for sequence classification, and whether a similar structure to the full-length tree would be observed. Analysis of *stevor* semiconserved N terminal regions based on sequences obtained from laboratory isolates 3D7, Dd2 and HB3 revealed a different tree structure with short branch lengths (Figure 3.10), better branch support and overall better statistical likelihood (log likelihood of -19,464) than the full length tree. The subgrouping especially for *stevorA* was significantly different to what was observed in the full-length (Figure 3.7) or HVR (Figure 3.9) trees. The *stevorA* sequences were reordered into different groups resulting in new combinations of subgroups. The *stevorB* groups were maintained suggesting that the *stevorB* are evolutionary constrained. This tree structure was more similar to the full-length tree (Figure 3.7) especially for *stevorB* groups with the major changes observed with *stevorA* groups.

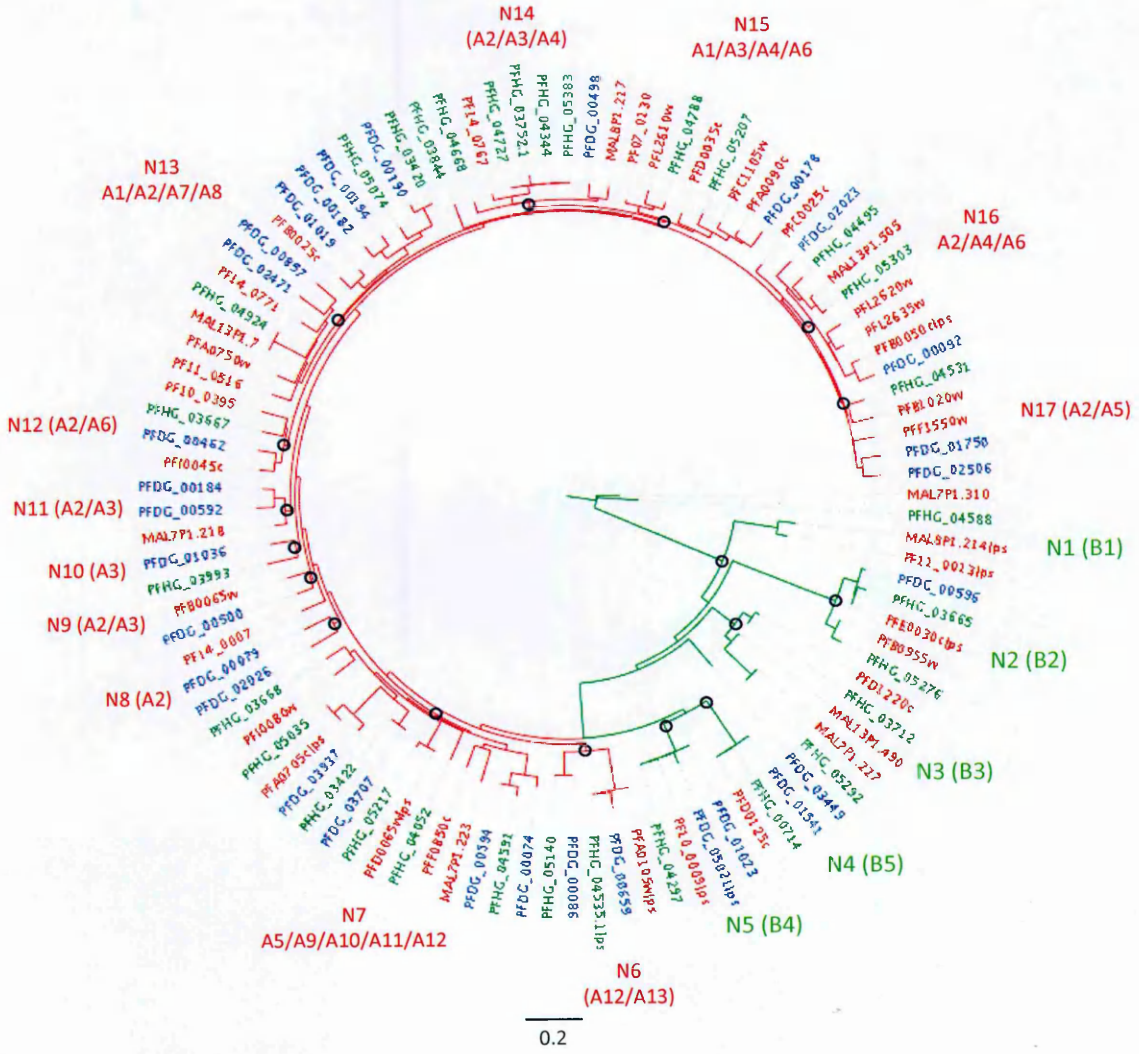

N14
(A2/A3/A4)
N15
A1/A3/A4/A6
N16
A2/A4/A6
N17 (A2/A5)
N1 (B1)
N2 (B2)
N3 (B3)
N4 (B5)
N5 (B4)
N6
(A12/A13)
N7
A5/A9/A10/A11/A12
N8 (A2)
N9 (A2/A3)
N10 (A3)
N11 (A2/A3)
N12 (A2/A6)
N13
A1/A2/A7/A8
0.2

***Figure 3.10**. Maximum likelihood tree of 102 N terminal stevor nucleotide sequences from laboratory isolates 3D7 (red), Dd2 (blue) and HB3 (green) showing the two larger stevor groups the major group A (red branches) and the minor group B (green branches). The larger groups further split into smaller groups. The clades are marked with a black open circle. The stevorA splits into 12 subgroups with most sequences reordered into different groups from the groups observed in the full-length tree resulting into new combination of groups while stevorB maintains the same 5 subgroups. The tree was constructed from MUSCLE-alignments of 102 stevor CDS nucleotide sequences (3D7=40, DD2=31 and HB3=31), using PhyML (in house), GTR model. The tree was viewed and edited using FigTree v1.4.0. Log-likelihood: - 19,464.*

3.4.3. Phylogenetic analysis of *rif* upstream regions in laboratory isolates.

As previously discussed, certain *var* genes are associated with specific upstream (ups) sequences. Phylogenetic analysis was carried on *pir* 5' UTRs to explore whether these ups groups are associated with specific *pir* groups. *Rif* UTRs, 1000 nucleotides upstream of the start codons, for 3D7 (184) and IT (138) *rif* genes were downloaded from GeneDB (http://www.genedb.org/genedb/malaria/). The IT genome still has gaps and a few *rif* Ups sequences had stretches of missing sequence data and were excluded from the analysis ending up with 120 IT *rif* Ups sequences.

The *rif* upstream regions clustered into 13 groups, 9 groups predominantly made up of *rifA* and 4 with mostly *rifB* sequences, although the separation between *rifA* and *rifB* groups was not as apparent as in the CDS region (Figure 3.11). The *rifB* grouped together with *rifB1* and *rifB2* mostly clustered in the same clades. A few *rifB3* did cluster with *rifA* in the same clades in adjacent branches. The conserved *rifA* groups *rifA1* and *rifA3* clustered on the ups tree together with other *rifA* ups sequences. *rifA1* grouped together with other *rifA* in the ups clade U9 while *rifA3* fell into U1. Most of the other *rifA* groups previously seen in the full-length tree were not maintained in the ups tree.

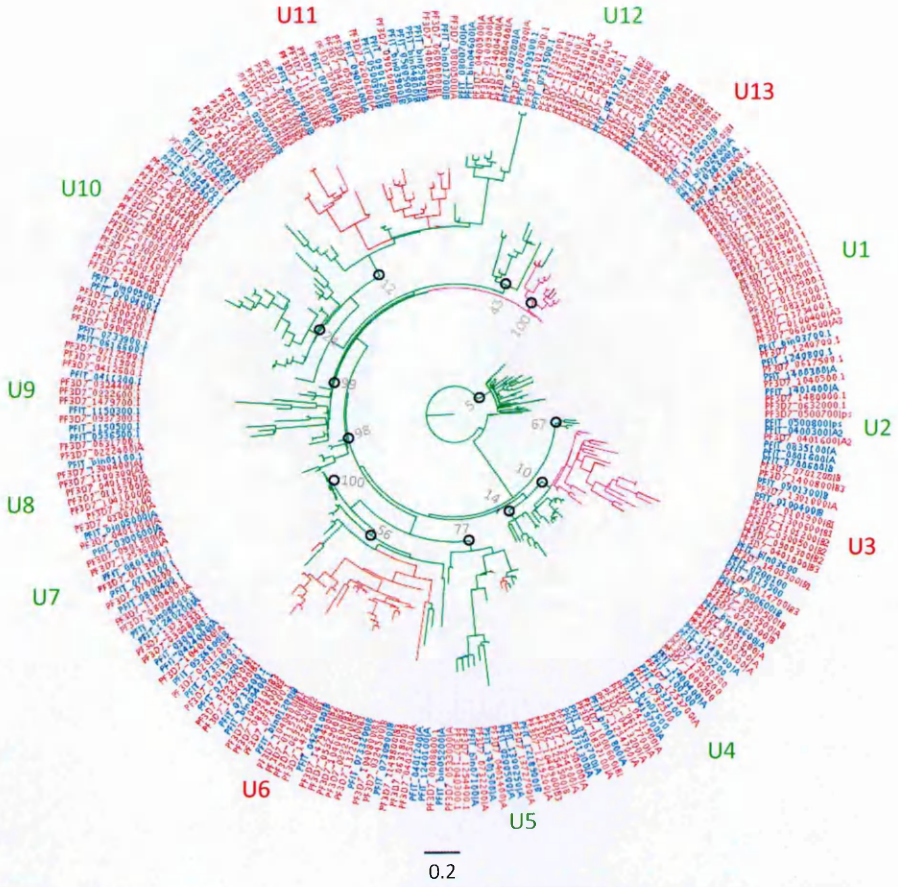


Figure 3.11. *Maximum likelihood tree of 3D7 (red) and IT (blue) rif upstream regions showing clustering into 13 Ups groups (U1 to U13), 9 predominantly rifA (green branches) and 4 predominantly B groups (B1 pink, B2 brown and B3 orange branches). The branch support for the 13 groups are indicated (bootstrap values out of 100 replicates). The tree was based on MUSCLE-alignments of 304 rif upstream sequences (3D7=184 and IT= 120) coding sequences, constructed using PhyML (run in house). Model: GTR. . Log-likelihood: -180837. The tree was edited using FigTree v1.4.0.*

3.4.4. Analysis of *stevor* upstream regions in laboratory lines 3D7 and IT.

Similarly, *stevor* Ups regions consisting of 1000 nucleotides upstream of the start codons were downloaded from GeneDB (http://www.genedb.org/genedb/malaria/) for 3D7 (42) and IT (39).

The *stevor* ups sequences grouped into 3 large clusters, U1, U2 and U9 that contained a majority of the sequences, and 7 smaller clusters with between 2 and 7 sequences.

In the *stevor* ups tree *stevorA* and *stevorB* groups did not segregate but instead grouped together and in some cases in adjacent branches (Figure 3.12). With the exception of U8 and U10 that contained only *stevorB* ups sequences most other *stevorB* mixed in with *stevorA* ups sequences on the tree.

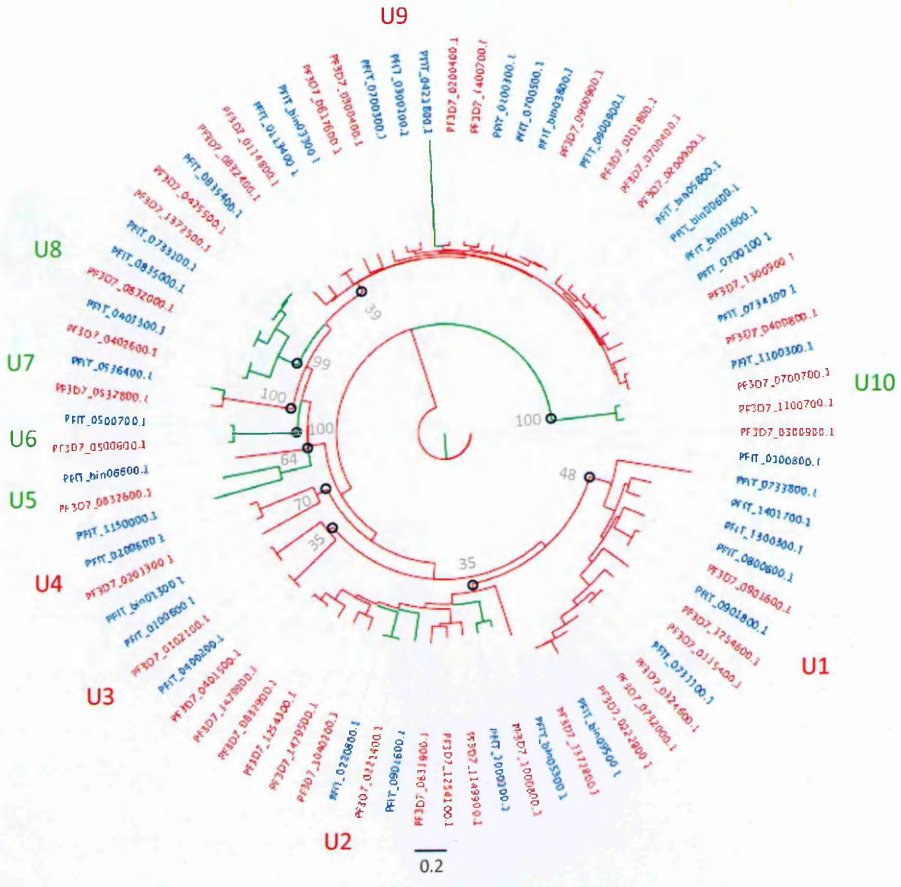


Figure 3.12. *Maximum likelihood tree of 3D7 (red) and IT (blue) stevor upstream regions showing 10 stevor Ups groups, 5 predominantly stevorA groups (red branches) and 5 predominantly stevorB groups (green branches). The branch support for groups are indicated (bootstrap values out of 100 replicates). The tree was based on MUSCLE-alignments of 81 rif (n=42) and stevor (n=39) coding sequences, constructed using PhyML (run in house). Model: GTR. . Log-likelihood: -39719. The tree was edited using FigTree v1.4.0.*

3.4.5. Conserved chromosomal locations of *pir* genes across parasite isolates and species.

Large multigene families in *Plasmodium* including *pir* genes are mainly located in the subtelomeric regions, but a few are found within the centromeric regions (Gardner et al., 2002, Pain et al., 2008, Carlton et al., 2008, Otto et al., 2014a, Otto et al., 2014b). The *pir* genes are found interspersed between the *var* genes, with the *vars* on the most subtelomeric regions, followed by a *rif* or a number of *rifs* and then *stevors* (Figure 3.13). In some chromosomes *var* genes appear singly or in a cluster in the centromeric regions. Some central *var* genes neighbor central *rif* and *stevor* genes (Figure 3.13). In the centromeric regions *rif* and *stevor* genes always appear next to a *var* gene. In 3D7 only 5 of the 14 chromosomes contain centrally located *vars* and *rifs,* chromosomes 4, 6, 7, 8 and 12. Of the 184 *rif* genes, 12 are centrally positioned. 6 of these are pseudogenes. Chromosome 7 has a cluster of central *var* genes and interspersed between them are 4 *rifs*, three of which are pseudogenes located on the reverse strand. All central *rif* genes in 3D7 belong to the *rifA* subgroup. Only one *stevor* gene (PF3D7_0617600) is centrally located on chromosome 6 next to a *var* gene and a *rif* pseudogene (Figure 3.13. A).
A similar overall chromosomal location is conserved in the genome of another laboratory isolate IT (Figure 3.13. B). Although not all genes have been assigned chromosome locations (43 *rif* and 9 *stevor* are currently in the bin, that is not assigned to any chromosome region), a majority of *pir* genes are located in the subtelomeric regions, with 7 *rif* genes centrally located. The central *rifs* are found in chromosomes 4, 6, 7, 8 and 12, as observed in 3D7. The central *rif* genes always appear next to *var* genes. Centrally located *stevor* genes were not observed in IT (Figure 3.13. B).

With the availability of high quality *P. reichenowi* genome, a close relative of P. falciparum, it was possible to analyse chromosomal locations of *pir* genes. *P. reichenowi* contains many more *pir* genes in comparison to *P. falciparum*, with 462 *rif* and 66 *stevor* compared to 184 *rif* and 42 *stevor* in 3D7. A good number of these genes have not been assigned to any chromosome (about 418 *rif* and 53 *stevor*), but the available chromosomal location data was used. The non-assigned *pir* genes were placed on an artificial chromosome zero (0) in descending order based on their gene IDs (Figure 3.13).

A comparison between the chimpanzee malaria parasite *P. reichenowi* and *P. falciparum* revealed that the chromosomal positioning of VSA coding genes was generally conserved. Most were located in the subtelomeric regions with a few found in the central regions of the chromosomes (Figure 3.13. C). Central *var* genes were observed in chromosomes 4, 6, 7, 8 and 12 as observed in 3D7 and IT. There were 9 central *rif* genes found in chromosomes 4, 7 and 12. Of these 4 were pseudogenes, one each in chromosome 4 and 12 and two in chromosome 7. All but one of the central *rif* were from the *rifA* group. The exception was the central *rif* PRCDC_0410600 on chromosome 4 that was from the *rifB1* subgroup. The central *rif* pseudogene on chromosome 12 (PRCDC_1218500) is a conserved gene with syntenic orthologs in 3D7 and IT (Figure 3.14 and Appendix table 9.3). Only one *P. reichenowi stevor* (PRCDC_0410700) was centrally located in chromosome 4. This *stevor* had no ortholog in *P. falciparum* but was closely related to the 3D7 stevor PF3D7_1372800 (MAL13P1.505) with a percent pairwise identity of 83% in 759 of 906 nucleotides.

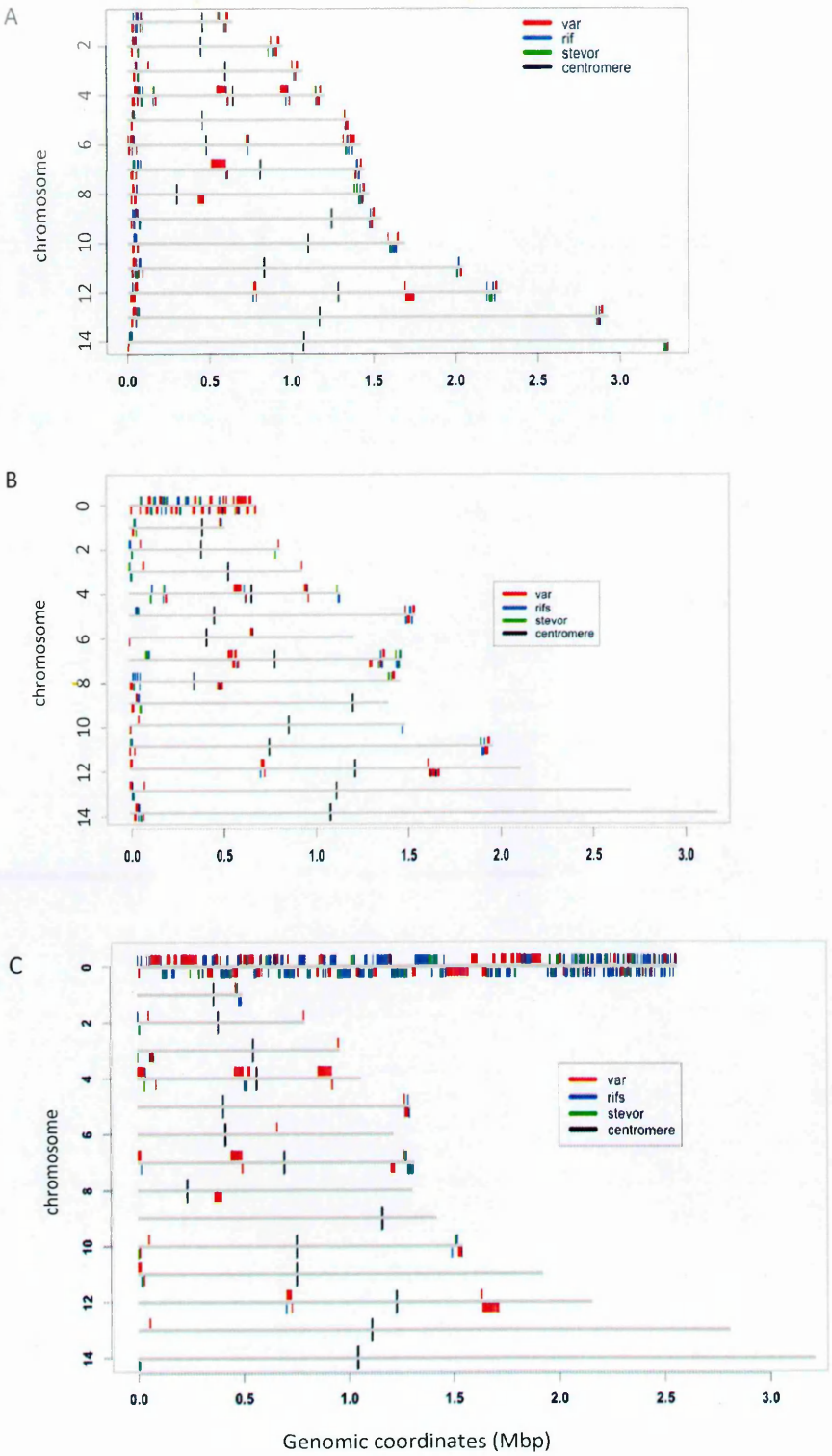

A
var
rif
stevor
centromere
chromosome
2
4
6
8
10
12
14
0.0
0.5
1.0
1.5
2.0
2.5
3.0
B
var
rifs
stevor
centromere
chromosome
0
2
4
6
8
10
12
14
0.0
0.5
1.0
1.5
2.0
2.5
3.0
C
var
rifs
stevor
centromere
chromosome
0
2
4
6
8
10
12
14
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Genomic coordinates (Mbp)

Figure 3.13. *Chromosomal organization of VSA-encoding genes, var (red), rif (blue) and stevor (green) in two P. falciparum parasite isolates 3D7 (A) and IT (B) (PF3D7 and PFIT) and the chimpanzee malaria P. reichenowi isolate PRCDC (version dated* 2014-09-03 from PlasmoDB) *(C). The 14 chromosomes are labeled on the Y-axis and the genomic coordinates on the X-axis in megabase pairs (Mbp). Chromosomes labeled 0 (zero) for IT and P. reichenowi contain VSAs that have not been assigned chromosomal locations and are still in the bin. The genes on chromosome 0 are ordered by name, in ascending order, along the chromosomes. The predicted centro meres are indicated in black bars.*

3.4.6. Centrally located *rif* genes are conserved in laboratory and wild isolates.

The *P. falciparum* central chromosomal regions are considered less "recombinogenic" as compared to the telomeric regions, thus centrally positioned genes are in turn expected to be more conserved. A group of *var* genes, UpsC *var* located near the centromeres are proposed to be conserved because of the reduced recombination events. As previously discussed, there are *rif* genes located next to the central *vars*. To analyse whether central *rifs* formed a conserved group, 28 central *rif* sequences from 3D7 (n=12), IT (n=7) and *P. reichenowi* (n=9) were extracted, aligned, and a phylogenetic tree constructed from the alignment. With the exception of a divergent 3D7 *rif* pseudogene (PF3D7_0617500/ PFF0847w) all the other *rif* genes fell into clusters with one or more closely related *rif*, within and across isolates (Figure 3.14). The high branch support and high log likelihood suggest that these centrally located *rifs* formed a conserved group. Six pairs of *rifs* from 3D7 and IT fell in the same clades (Figure 3.14) suggesting that these were closely related. A one-to-one pairwise comparison between the 3D7 and IT *rif* genes that fell in the same clades revealed that they were indeed orthologs (between 95% to 100% pairwise identity) (Appendix Table 9.3). In fact, four of these are syntenic orthologs (Table 3.4 and Appendix Table 9.3) (PlasmoDB (http://www.plasmodb.org/plasmo/home.jsp). Three clades with long branches but high bootstrap values were composed of pairs of sequences from 3D7 and *P. reichenowi* (Figure 3.14). The *rifs* in these clades were not orthologs based on results from pairwise sequence comparisons (below 70% pairwise identity). These *P. reichenowi* and *P. falciparum rif* genes might have shared ancestry but have diverged further after speciation of the parasites to different hosts. One clade with high branch support was

made up of *rif* sequences from chromosome 12 of 3D7, IT and P. reichenowi. The three central *rif* pseudogenes PF3D7_1219200, PFIT_1218900 and PRCDC_1218500 were positionally conserved in the three parasite genomes and had an overall pairwise identity of 97% (Table 3.4, Appendix Table 9.3 and Figure 3.15). The upstream regions of all the syntenic orthologs of centrally located *rifs* were conserved with identities between 90% and 98% over the full length of the sequences (Appendix table 9.3). This region in chromosome 12 is conserved in the 2 *P. falciparum* genomes 3D7 and IT and in *P. reichenowi* and contains a *var* gene, a *rif* pseudogene and

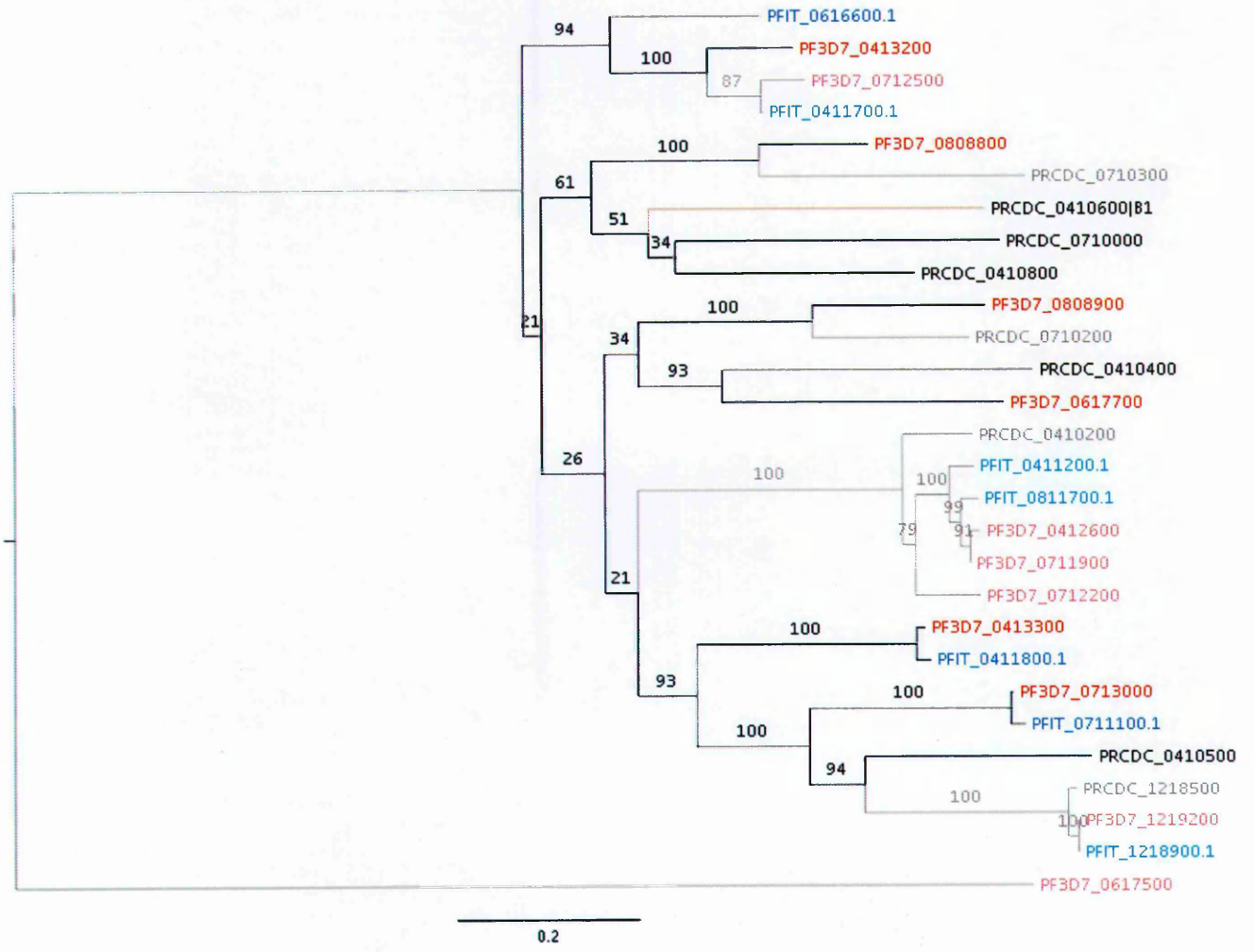

PFIT_0616600.1
PF3D7_0413200
PF3D7_0712500
PFIT_0411700.1
PF3D7_0808800
PRCDC_0710300
PRCDC_0410600|B1
PRCDC_0710000
PRCDC_0410800
PF3D7_0808900
PRCDC_0710200
PRCDC_0410400
PF3D7_0617700
PRCDC_0410200
PFIT_0411200.1
PFIT_0811700.1
PF3D7_0412600
PF3D7_0711900
PF3D7_0712200
PF3D7_0413300
PFIT_0411800.1
PF3D7_0713000
PFIT_0711100.1
PRCDC_0410500
PRCDC_1218500
PF3D7_1219200
PFIT_1218900.1
PF3D7_0617500
0.2

Figure 3.14. *Phylogenetic tree of the centrally located rif genes in P. falciparum isolates 3D7 (red), IT (blue) and the chimpanzee malaria parasite P. reichenowi (version dated* 2014-09-03 from PlasmoDB) *(black). The pseudogenes are on the gray branches with lighter shades of colour on the tip labels. The brown branch indicates the one centrally located rifB in P. reichenowi. The branches labeled indicate bootstrap values out of 100 replicates. The tree was based on MUSCLE alignments of 28 CDS rif sequences (3D7=12, IT=7 and PRCDC=9, constructed using PhyML with GTR model. 100 bootstrap replicates were performed. Log likelihood: -18408.*

The centrally located *rifs* formed a conserved group as discussed previously. To find out if these *rifs* were conserved in other parasite isolates BLAST searches were carried out using the 3D7 or IT representative as the query. Highly conserved orthologs were identified in the laboratory isolates DD2 and HB3 and in patient isolates from Kenya (Table 3.4). All central *rifs* had at least two orthologs from wild isolates and most had orthologs from other laboratory isolates, forming 12 groups of orthologous *rif* genes. Four of these groups had syntenic orthologs in 3D7 and IT, and in one case, PF3D7_1219200, the synteny was also conserved in *P. reichenowi*. Five of these orthologous groups were pseudogenes and they showed higher within group sequence identity on average as compared to the intact coding genes (Table 3.4).

The central 3D7 *stevor* PF3D7_0617600/ PFF0850c was used to search for orthologs in other laboratory isolates as well as wild isolates. There were no orthologs identified in the other three laboratory isolates DD3, HB3 and IT. Orthologs were found in three wild isolates, with an overall sequence identity of 98.3% (Figure 3.16).

Table 3.4. *Conserved centrally located rif genes and their orthologs from laboratory and patient isolates.*

Group	Lab Isolate	Gene ID	Old ID (3D7)	Pseudogene	Idnetitical nucleotides (% pairwise identity)	Patient isolate	Gene ID	% mean pairwise identity	Group % pairwise identity
1	3D7	PF3D7_0412600 (s)	PFD0620c	Y	Query	6816	PF8443_8_1_U561100	980/1023 (95%)	93.4
	IT	FIT_0411200 (s)			975/1023 (95%)	10936	PF8443_8_10_U489500	986/1023 (96%)	
	3D7	PF3D7_0711900	MAL7p1.43		1013/1074 (94%)	10936	PF8443_8_10_U489700	985/1073 (91%)	
	IT	PFIT_0811700			851/934 (91%)	10936	PF8443_8_10_U489800	959/1036 (92%)	
	HB3	PfHB3_BroadWash144100	N/A		1019/1074 (94%)	9626	PF8443_8_2_U283400	885/939 (94%)	
	HB3	PfHB3_BroadWash054500			971/1053 (92%)	11019	PF8443_8_3_U116300	970/1023 (94%)	
						11019	PF8443_8_3_U546400	787/816 (96%)	
						9215	PF8443_8_4_U515500	909/959 (94%)	
						9215	PF8443_8_4_U515600	853/905 (94%)	
						9775	PF8443_8_5_U058900	825/859 (96%)	
						10668	PF8443_8_8_U508400	828/870 (95%)	
						11014	PF8443_8_9_U052000	928/970 (95%)	
						SA075	PFSAO75_071600	978/1074 (91%)	
						SA075	PFSAO75_137600	951/1023 (92%)	
2	3D7	PF3D7_0808900	PF08_0104	N	Query	6816	PF8443_8_1_U073300	916/1000 (91%)	91.3
	HB3	PfHB3_BroadWash172400			912/999 (91%)	6816	PF8443_8_1_U547100	916/999 (91%)	
	IT	PFIT_1240200			921/999 (92%)	10936	PF8443_8_10_U548000	939/999 (93%)	
	IT	PFIT_bin08400			916/1005 (91%)	9626	PF8443_8_2_U143100	917/999 (91%)	
	DD2	PFDG_04323			928/999 (92%)	9626	PF8443_8_2_U513000	918/1002 (91%)	
						11019	PF8443_8_3_U164800	965/1020 (94%)	
						11019	PF8443_8_3_U477700	923/999 (92%)	
						9215	PF8443_8_4_U071400	876/961 (91%)	
						9215	PF8443_8_4_U195200	911/1001 (91%)	
						9775	PF8443_8_5_U165400	884/950 (93%)	
						9605	PF8443_8_6_U548500	923/1002 (92%)	
						9106	PF8443_8_7_U513600	908/999 (90%)	
						10668	PF8443_8_8_U293700	952/999 (95%)	
						11014	PF8443_8_9_U142300	951/999 (95%)	
						11014	PF8443_8_9_U162600	916/999 (91%)	
3	3D7	PF3D7_0713000 (s)	MAL7P1.57	N	Query	10936	PF8443_8_10_U132900	508/525 (96%)	97.4
	HB3	PfHB3_BroadWash144500			512/525 (97%)	10975	PF8443_8_12_U137000	924/963 (95%)	
	IT	PFIT_0711100 (s)			1021/1041 (98%)	9626	PF8443_8_2_U127100	1026/1041 (98%)	
						11019	PF8443_8_3_U123700	1010/1041 (97%)	
						9215	PF8443_8_4_U144300	1025/1041 (98%)	
						9775	PF8443_8_5_U139700	1009/1041 (96%)	
						9605	PF8443_8_6_U140200	1027/1041 (98%)	
						9106	PF8443_8_7_U060200	994/1041 (95%)	
						10668	PF8443_8_8_U135200	1022/1024 (99%)	
						11014	PF8443_8_9_U121800	1010/1041 (97%)	
						SA075	PFSAO75_137800	508/525 (96%)	
4	3D7	PF3D7_1219200 (s)	PFL0933w	Y	Query	10936	PF8443_8_10_U495100	625/629 (99%)	99.4
	IT	PFIT_1218900 (s)			704/704 (100%)	10975	PF8443_8_12_U312700	701/704 (99%)	
	PRCDC	PRCDC_1218500 (s)			688/704 (97%)	9626	PF8443_8_2_U300600	700/704 (99%)	
						11019	PF8443_8_3_U279800	640/644 (99%)	
						9215	PF8443_8_4_U327700	626/629 (99%)	
						9775	PF8443_8_5_U324600	701/704 (99%)	
						9106	PF8443_8_7_U276900	626/629 (99%)	
						10668	PF8443_8_8_U309400	625/629 (99%)	
						11014	PF8443_8_9_U278200	701/704 (99%)	
						SA075	PFSAO75_318800	701/704 (99%)	

	Lab Isolate	Gene ID	Old ID (3D7)	Pseudogene	% mean pairwise identity	Patient isolat	Gene ID	% mean pairwise identity	Group % pairwise identity
5	3D7	PF3D7_0413300 (s)	PFD0645w	N	Query	6816	PF8443_8_1_U061600	1031/1041 (99%)	96
	IT	PFIT_0411800 (s)			1017/1041 (97%)	10975	PF8443_8_12_U07020	846/860 (98%)	
	DD2	PFDG_03424			1037/1041 (99%)	9626	PF8443_8_2_U053400	1029/1041 (98%)	
	PRCDC	PRCDC_0410800*			765/1049 (72%)*	11019	PF8443_8_3_U052600	1015/1041 (97%)	
						9215	PF8443_8_4_U059600	945/1041 (90%)	
						9775	PF8443_8_5_U059100	1037/1041 (99%)	
						10668	PF8443_8_8_U056700	945/1041 (90%)	
						11014	PF8443_8_9_U052300	1037/1041 (99%)	
						SA075	PFSAO75_061500	1016/1041 (97%)	
6	3D7	PF3D7_0413200	PFD0640c	N	Query	6816	PF8443_8_1_U061500	1050/1134 (92%)	94.1
						9626	PF8443_8_2_U053300	1131/1131 (100%)	
						11019	PF8443_8_3_U052500	1103/1131 (97%)	
						9775	PF8443_8_5_U059000	889/939 (94%)	
						9605	PF8443_8_6_U059100	1102/1131 (97%	
						10668	PF8443_8_8_U056600	1105/1131 (97%)	
						SA075	PFSAO75_061400	1021/1131 (90%)	
7	3D7	PF3D7_0808800	PF08_0105	N	Query	9215	PF8443_8_4_U171000	998/1070 (93%)	91.4
	DD2	PFDG_04322			1008/1114 (90%)	10668	PF8443_8_8_U293600	1022/1113 (91%)	
						6816	PF8443_8_1_U560100	1020/1113 (91%)	
						9626	PF8443_8_2_U143200	879/891 (98%)	
						11014	PF8443_8_9_U502600	1014/1116 (90%)	
						6816	PF8443_8_1_U547200	964/1038 (92%)	
						11019	PF8443_8_3_U568200	923/976 (94%)	
8	IT	PFIT_0616600		N	Query	10936	PF8443_8_10_U110500	1048/1148 (91%)	94
	DD2	PFDG_02423			663/725 (91%)	9215	PF8443_8_4_U134400	1137/1147 (99%)	
						9775	PF8443_8_5_U130900	1137/1147 (99%)	
						9106	PF8443_8_7_U112500	1049/1148 (91%)	
						SA075	PFSAO75_PN18	1059/1069 (99%	
9	3D7	PF3D7_0712500	MAL7P1.52	Y	Query	9775	PF8443_8_5_U139500	1107/1137 (97%)	98.4
	IT	PFIT_0411700			995/997 (99%)	9626	PF8443_8_2_U126900	1089/1137 (95%)	
	DD2	PFDG_00027			1106/1139 (97%)	6816	PF8443_8_1_U141200	1043/1065 (97%)	
10	3D7	PF3D7_0617500	PFF0847w	Y	Query	9106	PF8443_8_7_U100400	886/892 (99%)	99.5
						SA075	PFSAO75_071100	582/586 (99%)	
						9775	PF8443_8_5_U530200	582/586 (99%)	
11	3D7	PF3D7_0617700	PFF0855c	N	Query	9106	PF8443_8_7_U100500	906/1003 (90%)	92.6
						SA075	PFSAO75_071300	717/752 (95%)	
12	3D7	PF3D7_0712200	MAL7P1.47	Y	Query	9775	PF8443_8_5_U139300	1014/1017 (99%)	94.9
						11014	PF8443_8_9_U435200	944/1018 (92%)	

Yellow highlights indicate the centrally located rifs.

The reference gene was used to search for orthologs using BLAST search in Geneious (version 6.1.5).

s indicates syntenic orthologs

** non-orthologs because of the low pairwise sequence identity*

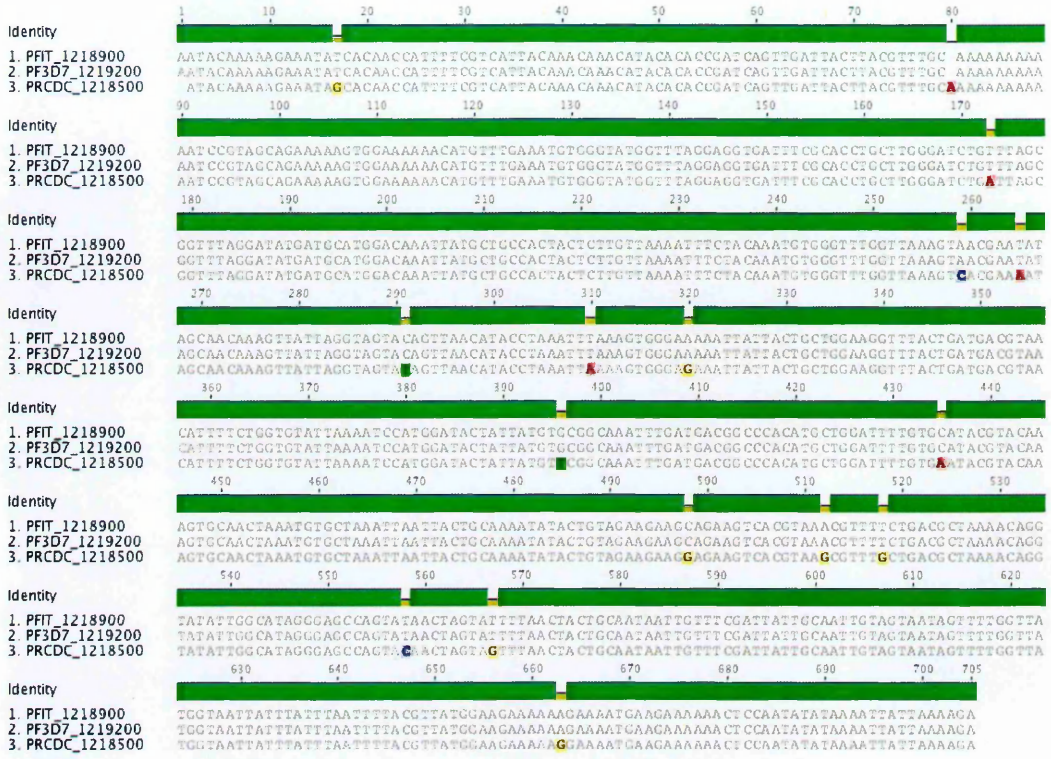


Figure 3.15. *Nucleotide sequence alignment of a conserved central rif pseudogene (exon 2 only) in P. falciparum isolates 3D7 and IT and in P. reichenowi isolate PRCDC. Above the sequences is the identity score, the mean pairwise identity of all the pairs in the column; green: 100% identity, brown-green: 30-99%, and red: less than 30% identity. The overall pairwise identity of the four sequences was 97%.*

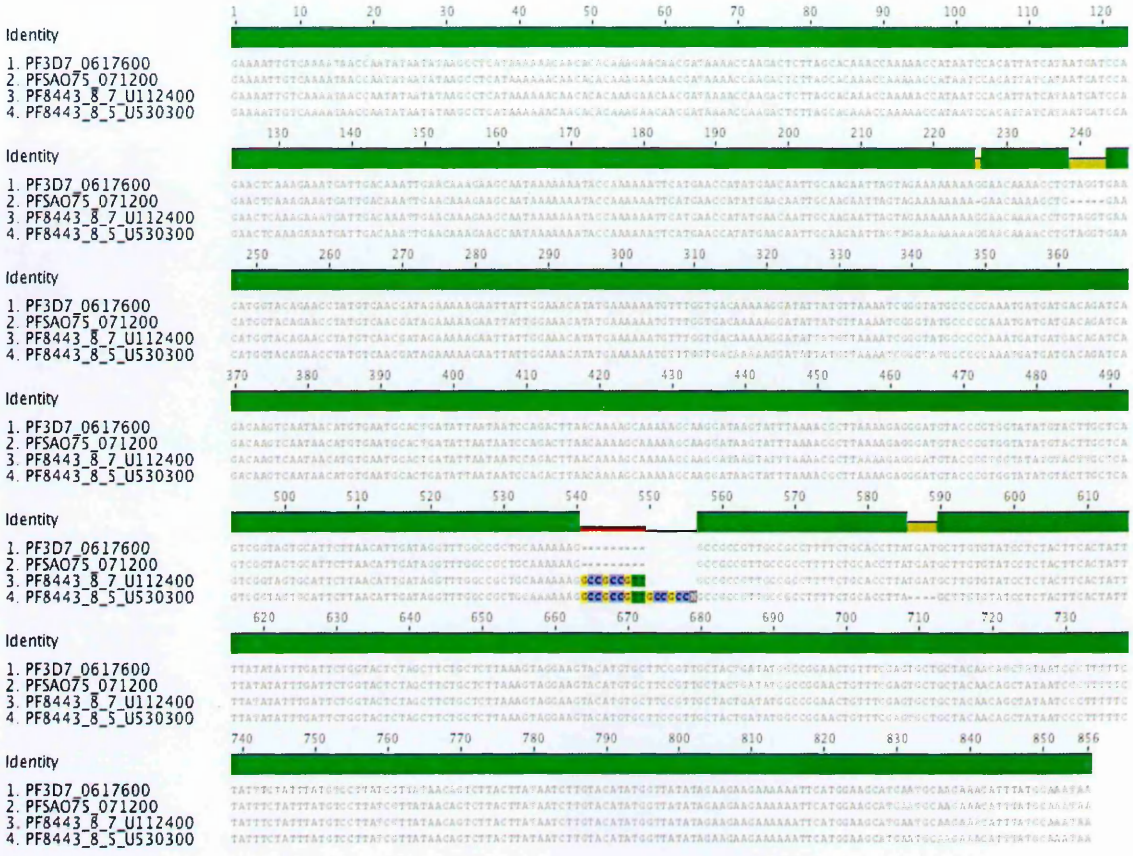


Figure 3.16. *Nucleotide sequence alignment of a conserved central stevor gene in P. falciparum laboratory isolate 3D7 and three wild isolates. Above the sequences is the identity score, the mean pairwise identity of all the pairs in the column; green: 100% identity, brown-green: 30-99%, and red: less than 30% identity. The overall pairwise identity of the three sequences was 98.3%.*

3.4.7. *Rif* and *stevor* sequence length analysis.

On average, *rifA* are longer than *rifB*, 1227 and 1115 basepairs (bp) for *rifA* and *rifB* respectively from 3D7 genome. This is mainly due to the 25 amino acid insertion in *rifA* (Figure 3.1). Distribution of all the *rif* CDS lengths appears to follow a normal distribution, a few very short or very long sequences with a majority of the sequences longer than 900 nucleotides and shorter than 1150 nucleotides. When separated by type, sequences separate by length with *rifA* having 3 length groups, short (less than 1000 nulceotides), medium long (1000 to 1100 nucleotides) and long (longer than 1100 nucleotides). Two length groups were observed for *rifB*, short (less than 1000 nucleotides long) and long (longer than 1000 nucleotides) (Figure 3.17). From phylogenetic analysis of CDS from 3D7 it was evident that *rif* sequences clustered by length, especially the longer *rifs* that were mostly found to form their own clusters (Figure 3.18). Multiple alignments of the different groups by length revealed that the longer sequences were more conserved than the shorter sequences, albeit only slightly. For *rifA* the mean pairwise identity of the sequences longer than 1100 nucleotides was 62.9% (n=53) as compared to 57.5% for the sequences between 1000-1100 nucleotides (n=44) or 55.9% for the sequences shorter than 1000 nucleotides (n=33). The overall pairwise identity for *rifA* from 3D7 genome was 54%. Similary for *rifB*, the longer sequences were more conserved than the short ones. The short *rifB* were all from *rifB3* subgroup and had an overall pairwise identity of 60.3 (n=16) as compared to 69.9% in the longer *rifB*. The overall pairwise identity for all *rifB* from 3D7 genome was 61.6%.

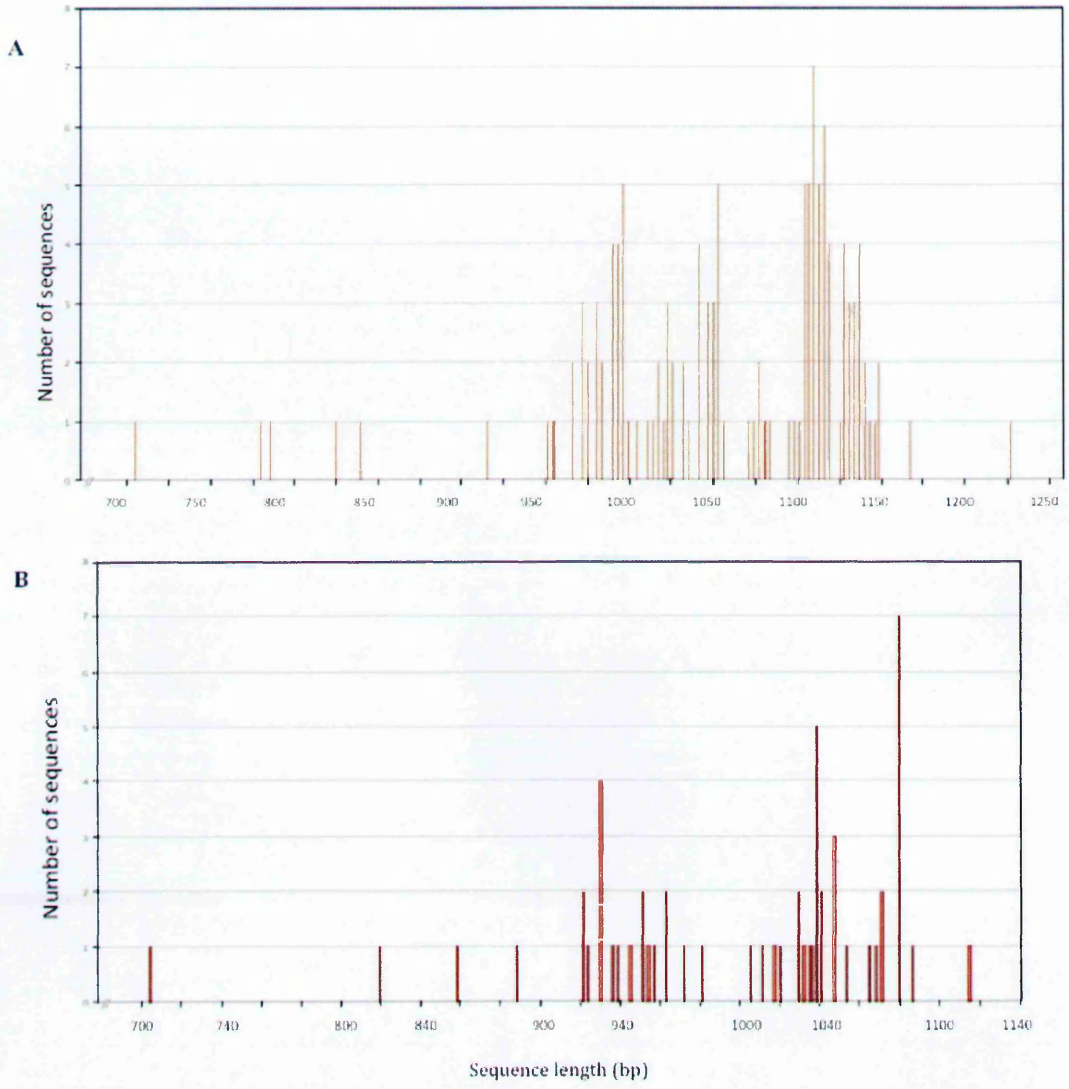


Figure 3.17*. Length distribution of 129 rifA (A) and 48 rifB (B) intact full length CDS sequences from 3D7 genome showing the three length groups for rifA, short, medium and long (below 1000, 1000 to 1100, and above 1100 nucleotides respectively) and two length groups for rifB, short and long (below 1000 and above 1000 nucleotides repsectively).*

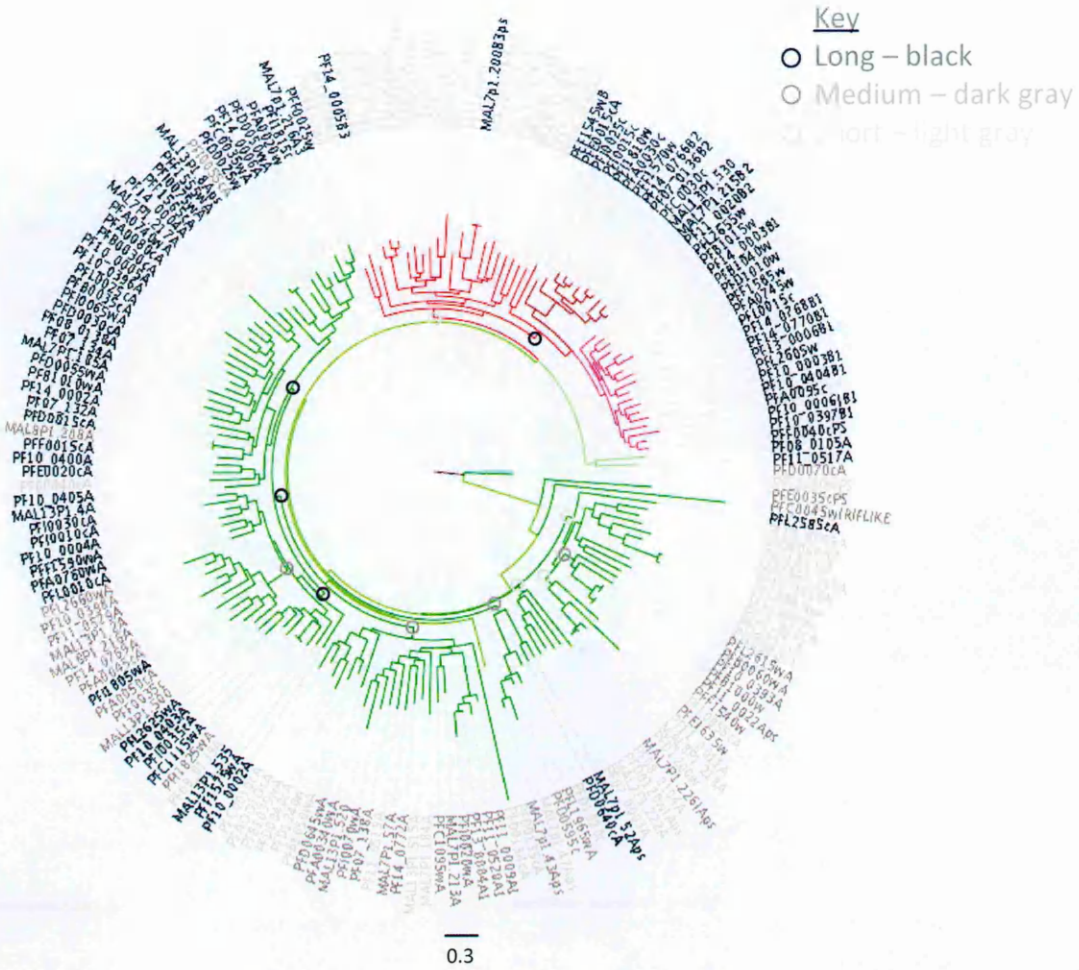
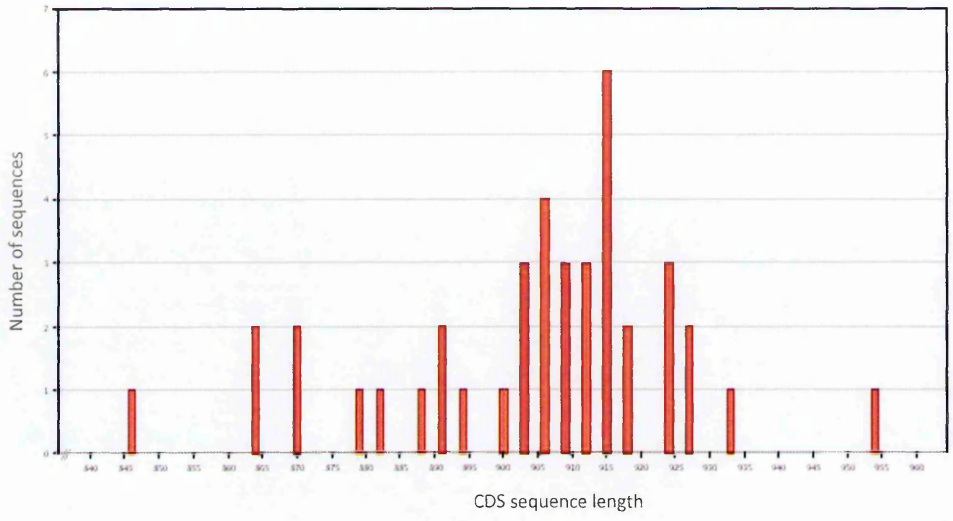


Figure 3.18. *Maximum likelihood tree of the 3D7 rif CDS gene family from showing clustering of rif genes by length. Tip labels are coloured by length, black for long sequences (rifA >1100, and rifB >1000 nucleotides), darker gray for medium long rif (>1000<1100 nucleotides only rifA) and light gray for short rif (<1000 nucleotides). Branches are coloured based on rif major groups, A (green) and B subgroups, B1 (pink), B2 (brown) and B3 (orange) subgroups (green). The tree was constructed from MUSCLE-alignments of 184 rif CDS, using PhyML (run on the ATGC bioinformatics platform South of France: http://www.atgc-montpellier.fr), GTR model, viewed and edited using FigTree v1.4.0. Log-likelihood: - 112112.*

Two sequence length groups were observed for *stevor* sequences from 3D7, short sequences less than 900 nucleotides and long sequences longer than 900 nucleotides (Figure 3.19. I). Similar to the observation in *rif* sequences, longer *stevor* sequences were slightly more conserved with 69.5% mean pairwise identity (n=28) than the shorter sequences with 62.3% mean pairwise identity (n=12). Overall stevor pairwise identity was 65.5% (n=42). Phylogenetic analysis did not show any clustering of stevor sequences by length (Figure 3.19. II).

I



II

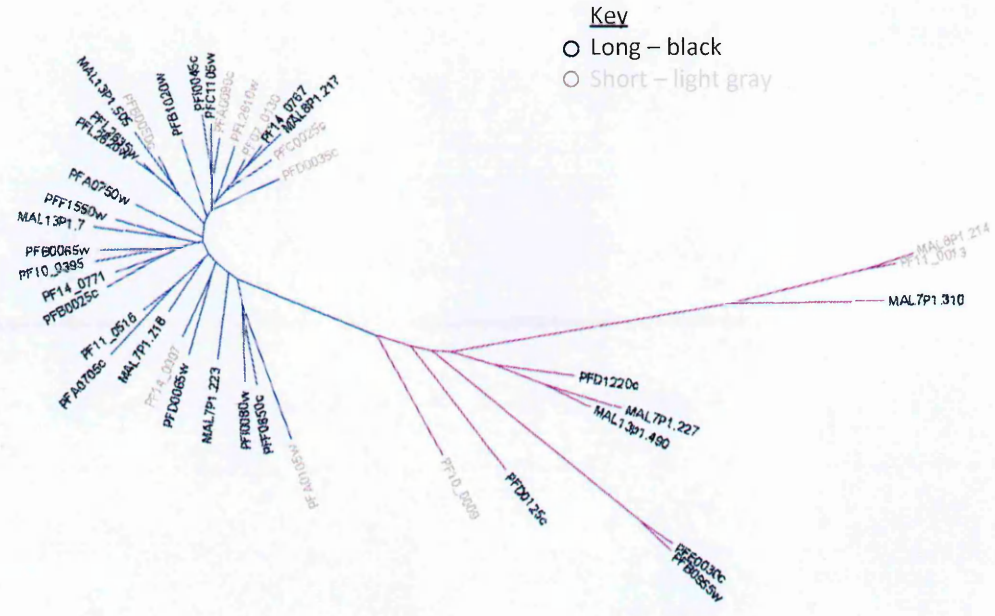


Figure 3.19. *Analysis of stevor sequences by length (I) Length distribution of 40 intact full length stevor CDS sequences from 3D7 genome showing distribution of stevor into two length groups (short and long, <900 and >900 nucleotides respectively. (II) A maximum likelihood tree of 40 intact full length stevor CDS sequences showing the main groups stevorA (blue branches) and stevorB (purple branches).*

3.4.8. Conserved *rif* and *stevor* genes

Despite the high diversity in *pir* genes especially the *rif* genes, there are highly conserved groups and individual genes. The major *rif* groups A and B, and B1, B2 and B3 subgroups have been described (Joannin et al., 2008) while for *stevor,* the major and minor groups were described in laboratory isolates (Schreiber et al., 2008). Some of the highly conserved *rifs* have been described in literature, including the ancient *rif* found as a single copy in all genomes *rifA2* (PF3D7_0401600/PFD0070c), the UpsA *var*-linked *rif*, (Wang et al., 2009), and the strain transcendent *rif* and *stevor* observed in laboratory and field isolates (Claessens et al., 2011).

With the availability of more *pir* repertoires from whole genome sequencing projects it was possible to identify orthologs from more laboratory and wild isolates from different geographical locations. In addition, the availability of the *P. reichenowi* whole genome sequence data enabled identification of ancient conserved *pir* genes.

To identify conserved *pir* genes in other laboratory and wild isolates 3D7 *pirs* were used as queries in BLAST searches (run within Geneious) against a database of *pir* sequences. The databases contained *pir* repertoires from *P. falciparum* and *P. reichenowi.* The *P. falciparum pir* genes were obtained from 4 laboratory isolates (3D7, IT, DD2 and HB3) and 13 wild isolates from Kenya (12 Kilifi and 1 Kisumu isolate). Additional incomplete *rif* repertoires were obtained from published data, field isolates from Brazil (Albrecht et al., 2006) and Gabon (Abdel-Latif et al., 2002). *P reichenowi pir* repertoires were obtained from the recently published PRCDC genome sequenced by the Sanger Institute (Otto et al., 2014b). Genes were considered to be orthologs if the nucleotide pairwise identities were above 90% over greater than half the length of the query and hit. A *pir*

gene was considered to be strain-transcendent if it had orthologs in two different isolates, that is, the gene was found in at least three different isolates (Claessens et al., 2011).

Using this approach highly conserved *pir* genes were identified. 35 *rif* and 29 *stevor* genes from 3D7 had at least two orthologs, either from two different laboratory or wild isolate, or one each from a laboratory and wild isolate (Table 3.5 below and Appendix Table 9.3 and 9.4). This threshold was intentionally lowered because many isolates, especially the wild isolates have incomplete genome annotations. The implication is that there may be more conserved *pir* genes that have not been annotated in some genomes and increasing the threshold would result in identification of only a few conserved *pir* genes. Certain chromosomal regions of the *P. falciparum* genome are highly conserved. Some of the conserved *pirs* were found within these highly conserved regions.

Of the 35 conserved *rif* genes, 23 were A type and 12 B type. Of these 13 were pseudogenes, 12 *rifA* pseudogenes and 3 *rifB* pseudogenes. 10 conserved *rif* genes were syntenic orthologs in 3D7 and IT genomes. Of the 35 conserved *rif* 2 had orthologs in P reichenowi including a highly conserved syntenic orthologous pseudogene on chromosome 12 (PF3D7_1219200). 3 other conserved *rif* had close relatives in *P reichenowi* (above 80% sequence identity).

Many 3D7 stevor genes had at least two orthologs either from laboratory or wild isolates. 29 of the 42 *stevors* were conserved in at least one other isolate (Appendix Table 9.4). Of these 21 were from the major group *stevorA* and 9 were pseudogenes. 9 of the conserved stevor shared synteny with orthologs from IT. 10 strain transcendent *stevor* genes,

conserved in at least 2 laboratory and 2 wild isolates, were observed (Table 3.5). 7 of the strain transcendent *stevor* had syntenic orthologs in IT. 2 of the conserved *stevorA* pseudogenes had orthologs in *P. reichenowi*.

Table 3.5. *Strain transcendent rif and stevor genes.*

Gene	Type	3D7	IT	HB3	DD2	Conserved in Brazilian isolates	Conserved Kenyan isolates (Kilifi)	Conserved Kenyan isolates (Kisumu)	Conserved in P. reichenowi
rif	A	PF3D7_0808900	PFIT_1240200	PfHB3_BroadWash172400	PFDG_04323	N	Y	N	N
rif	A	PF3D7_0412600 (s)	PFIT_0411200 (s)	PfHB3_BroadWash144100	N	N	Y	Y	N
rif	B ps	PF3D7_0114700 (s)	PFIT_0113500 (s)	PfHB3_BroadWash537700	PFDG_03938	Y	Y	Y	Y
rif	A ps	PF3D7_0713000 (s)	PFIT_0711100 (s)	PfHB3_BroadWash144500	N	N	Y	Y	N
rif	A ps	PF3D7_1101400	PFIT_1100600	N	PFDG_04006	N	Y	Y	N
rif	A	PF3D7_0402500 (s)	PFIT_0401200 (s)	PfHB3_BroadWash070900	N	Y	Y	Y	N
rif	A2	PF3D7_1219200 (s)	PFIT_1218900 (s)	N	N	N	Y	Y	N
rif	A ps	PF3D7_0413300 (s)	PFIT_0411800 (s)	N	PFDG_03424	N	Y	Y	N
rif	B ps	PF3D7_0401600	PFIT_0400300.2	PfHB3_BroadWash518600	PFDG_04771	N	Y	Y	N
rif	B3	PF3D7_0421500	PFIT_0419700	PfHB3_BroadWash075100	N	N	N	N	N
rif	B3	PF3D7_0731900 (s)	PFIT_0730900 (s)	PfHB3_BroadWash532400	PFDG_02022	N	Y	Y	N
rif	A ps	PF3D7_0532900	N	PfHB3_BroadWash536000	PFDG_02919	Y	Y	Y	N
rif	A	PF3D7_1372600	PFIT_0733200	PfHB3_BroadWash519800	PFDG_00493	N	Y	Y	N
rif	A ps	PF3D7_0533000 (s)	PFIT_0536700 (s)	PfHB3_BroadWash536100	N	N	Y	Y	N
rif	A	PF3D7_0808800	PFDG_04322	N	N	N	Y	N	N
rif	A	PF3D7_1240700	PFIT_1240800	PfHB3_BroadWash054500	N	N	Y	Y	N
rif	A	PFIT_0616600	N	N	PFDG_02423	N	Y	Y	N
rif	A ps	PF3D7_1253700	PFIT_0801500	PfHB3_BroadWash515900	N	Y	Y	N	N
rif	A	PF3D7_0901100	N	PfHB3_BroadWash534600	N	N	Y	N	Y ps
rif	B1	PF3D7_0402700 (s)	PFIT_0401400 (s)	N	N	N	Y	Y	N
rif	A	PF3D7_0712500	PFIT_0411700	N	PFDG_00027	N	Y	Y	N
rif	B1	PF3D7_1000600	N	PfHB3_BroadWash262300	N	N	Y	Y	N
stevor	B	PF3D7_0402600 (s)	PFIT_0401300 (s)	PfHB3_BroadWash071000	PFDG_01541	N	Y	Y	N
stevor	B	PF3D7_0832000 (s)	PFIT_0835000 (s)	PfHB3_BroadWash548600	N	N	Y	Y	N
stevor	B ps	PF3D7_1000800 (s)	PFIT_1000100 (s)	PfHB3_BroadWash262600	PFDG_0102	N	Y	Y	N
stevor	A ps	PF3D7_0102100 (s)	PFIT_0100600 (s)	PfHB3_BroadWash529200	N	Y	Y	N	N
stevor	B	PF3D7_0221400 (s)	PFIT_0220800 (s)	N	N	Y	Y	Y	N
stevor	A	PF3D7_0222800	PFIT_0700300	N	N	N	Y	Y	N
stevor	A	PF3D7_0732000	PFIT_0900900	N	N	N	Y	Y	N
stevor	A ps	PF3D7_0114600 (s)	PFIT_0113400 (s)	N	PFDG_03937	N	N	Y	Y ps
stevor	A	PF3D7_1300900	PFIT_0731100	PfHB3_BroadWash534700	N	N	Y	N	N
stevor	A	PF3D7_1372800	MAL13P1.505	PfHB3_BroadWash549600	N	N	Y	N	N
stevor	B ps	PF3D7_1372500 (s)	PFIT_0733100 (s)	N	N	N	N	Y	N
stevor	A	PF3D7_1479500	N	PfHB3_BroadWash531600	N	N	Y	N	N
stevor	A ps	PF3D7_0401500 (s)	PFIT_0400200 (s)	N	N	Y	N	Y	N
stevor	A	PF3D7_0500600 (s)	PFIT_0500700 (s)	N	N	N	N	Y	N
stevor	B	PF3D7_0200400	N	N	PFDG_01019	N	Y	N	N
stevor	A ps	PF3D7_0532800	N	N	N	N	N	Y	Y ps

Strain transcendent pirs defined as genes with orthologs (at above 90% pairwise identity over half the sequence length) in at least three different parasite isolates.

Yellow highlights indicate the centrally located rifs.

The reference gene was used to search for orthologs using BLAST search in Geneious (version 6.1.5).

s: syntenic orthologs

Y: Ortholog identified in wild isolates

N: No ortholog identified

3.5. Discussion.

A characteristic feature of *Plasmodium* parasites is the presence of large multigene families that encode VSA. These VSA are thought to play a role in host-parasite interactions, specifically immune evasion. In *P. falciparum* the major VSA, PfEMP1 encoded by *var* genes has been ascribed this role. The *var* genes are a more recent gene family and are only found in Laverania parasites with a homologous family *sicavar* in *P. knowlesi*. Therefore other genes encoding VSAs are probably key in mediating antigenic variation of the IE and may also play a role in cytoadherence. The *pir* multigene family one of the largest multigene families found in all *Plasmodium* species is a likely candidate. The *P. falciparum pir* gene families, *rif* and *stevor* encode VSAs that are coexpressed with PfEMP1 on the surface of IE. Preexisting data on repertoire diversity was limited to the WGS data available. With the availability of more high quality WGS data we sought to characterize sequence diversity in the *pir* gene families of *P. falciparum*.

Generally *rifA* and *rifB* are mainly differentiated by the presence of a 75 nucleotide insertion in the *rifA* but additional key sequence features differ between the two *rif* groups. It was interesting to observe certain genes from *rifA* clustering with *rifB* in the N terminal tree. The N terminal region contains the insert in *rifA* and therefore it was unexpected to observe mixing of *rifA* and *rifB* clusters based on the N terminal region. This observation indicates that some *rifA* and *rifB* are recombining more freely, and generate sequences that share both *rifA* and *rifB* features as a means of sequence diversification. The high sequence diversity accumulated in the HVR is suggestive of

immune diversifying selection pressure as well as higher rates of recombination within these regions. Although a lot more group mixing was observed in the tree based on HVRs it was surprising to see conserved groups maintaining cluster structures in the trees based only on the HVR. The *rif* subgroups that were structurally maintained in the HVR tree included *rifB1* and *rifB2*, the *UpsA rif* genes *rifA1* and *rifA3,* and the conserved ancient *rifA2* that always formed a separated cluster from all other *rif* in all trees. These groups formed conserved clades in the HVR tree suggesting restricted recombination that occurs only within group. On the other hand the diverse *rifB3* sequences mixed freely with some *rifA* on the tree. Many *rifA* groups were reorganized suggesting that there is little restriction to recombine among the non-*UpsA rifA* and *rifB3* groups in the HVRs.

Overall, phylogenetic trees obtained from the more conserved N terminal regions or the highly variable regions (HVR) had different fine structures from those observed in the full-length trees. The larger groups were generally maintained but smaller clades were shifted around the tree with certain genes moving from one subgroup to another. This is probably due to the different selection pressures that are imposed on the different regions of the sequences resulting in different evolutionary rates. For this reason it is difficult to obtain a clear picture of overall diversity in these genes without considering the full-length sequences. The limitation of clustering full length sequences is the inability to distinguish between the effects of diversifying selection pressure acting on the hypervariable regions and stabilizing selection pressure acting on the conserved and probably functional regions of the genes. It is therefore important for future analysis to

use a two-step approach initially clustering full length genes then clustering the different regions separately.

Phylogenetic analysis of *pir* upstream regions in two laboratory isolates revealed subgrouping of *pirs* that was different from the groupings observed with the coding sequences. The ups regions of the conserved *rif* groups rifA1, A2 and A3, were also conserved and clustered together. The ups regions of *rifB1* and *rifB2* were also mostly conserved and clustered together. A major limitation of the analysis on *pir* upstream regions was that it was done for only two laboratory isolates, 3D7 and IT, of which IT was missing sequence data because of incomplete WGS data. (Of the 120 *rif* ups sequences retrieved 26 had long stretches of missing sequence data.) For stevor 4 of the 39 ups sequences retrieved 4 had missing segments and one was truncated. A more comprehensive analysis of the *pir* upstream regions from laboratory and wild isolates will be useful to elucidate sequence clustering based on these regions.

A major limitation with this analysis is the inherent difficulty of clustering hypervariable regions or regions filled with sequence mosaics using traditional phylogenetic methods. In addition, these hypervariable regions may be undergoing recombination with small sequence blocks shuffling between sequences resulting in mosaics in the sequences. These traditional phylogenetic tree construction approaches generally assume lack of recombination between sequences and evolution via accumulation of point mutations. Thus functional groupings that may be dependent on smaller blocks of sequences or particular motifs within the HVRs may be lost using multiple alignments and tree

building. This limitation is evidenced by the long branches and low statistical support seen in the HVR trees (Figure 3.5 and 3.9) The preferred approach using non-alignment based network analysis was not possible within the scope of this PhD, but should be utilised in future to derive sequence subgroups.

Many *rif* and *stevor* sequences from 3D7 had orthologs and very closely related sequences from patient isolates from Kenya. In addition orthologs of 3D7 *rif* and *stevor* were more readily identified in genomes of patient isolates from Kenya than from other laboratory isolates IT, HB3 and DD2. This was despite the fact that the genomes from wild isolates were mostly incomplete with only partial *pir* repertoires; therefore many more orthologs are likely to be identified once the complete genomes are available. This observation of close orthologs is in line with the fact that parental line NF54 from which 3D7 originated is from Africa and therefore is more closely related to patient isolates from Kenya than from the other laboratory isolates from different parts of the world.

There are stretches of conserved chromosomal regions in 3D7 and IT genomes that contained conserved *pir* genes. The centromeric regions as expected did contain a number of *rif* and one *stevor* that made up a conserved group. This group was made up of both coding and pseudogenic sequences. Highly conserved orthologs of these central *pirs* were identified from other laboratory isolates (DD2 and HB3) and wild isolates as well as the related non-human malaria parasite *P. reichenowi*. Although having common ancestry the *P. falciparum* and *P. reichenowi* central *rif* genes appear to have undergone divergent evolution after speciation of the two parasites as could be seen by low sequence identity scores (below 90% pairwise identity). One exception is the highly conserved strain-

transcendent central *rif* pseudogene PF3D7_1219200/ PFL0933w with syntenic orthologs in IT and *P. reichenowi*, and also found in patient isolates from different geographical locations in Kenya (Kilifi and Kisumu). This *rif* pseudogene together with a conserved central *var* (PF3D7_1219300) are located in a highly stable region of chromosome 12 that is conserved in 3D7, IT and *P reichenowi*. In chromosome 1 and 5 several conserved *pirs* and a conserved *var* (PF3D7_0114400 in chromosome 1 and PF3D7_0533100 in chromosome 5) were observed. These regions are not centromeric, but are located in subtelomeric regions, which are thought to be highly recombinogenic, making the conservation even more peculiar. Another curious observation was the syntenic *pir* genes in two different chromosomes, chromosome 13 of 3D7 and right arm of chromosome 7 in IT. These two chromosomes contained the conserved syntenic *rif* (PF3D7_1372600 and PFIT_0733200) and *stevor* pseudogene (PF3D7_1372500 and PFIT_0733100) orthologs. Despite the high diversity of *pir* genes there appears to be conservation of repertoires in isolates from different geographical locations. For both gene families the subgrouping into A and B groups is conserved in all isolates. Recently, analysis of *pir* genes from the rodent malaria parasite *P. chabaudi cir* revealed a similar structure with A and B subfamilies. The CIR A family was larger and more diverse and was characterized by an insertion with similarities to the RIFIN-A insertion (Lawton et al., 2012). This gene structure conservation observed in a distantly related parasite might suggest that the *rifA* and *rifB* split occurred very early in evolution of these gene families, and might have been due to functional divergence.

Closely related *pirs* that are positionally conserved have been observed in *P. falciparum* and *P. reichenowi* suggesting that there are constraints to maintain sequence homology in certain regions of the chromosome. This could be either due to presence of important genes for parasite survival that are under stabilizing selection pressure or these regions function as homology blocks to facilitate recombination.

Strain transcending *rif* and stevor *genes* have been identified in laboratory and patient parasite isolates, and in the related non-human malaria parasite *P reichenowi.* These groups of *pirs* might be functionally constrained and more directed experiments might help elucidate their role in host parasite interactions.

CHAPTER 4

4. Analysing *rif* and *stevor* transcription profiles by amplicon cloning and capillary sequencing.

4.1. Introduction.

There is limited data on *rif* and *stevor* transcription in wild isolates. Studies based on laboratory isolates show VSA transcription to occur sequentially in overlapping windows during the parasite's IDC, with *var* transcription occurring at the early ring stage, followed by *rif* during the late ring early trophozoite stage, followed by *stevor* during the trophozoite stage (Kyes et al., 2000) (Kaviratne et al., 2002). Using northern blotting, Kyes and group showed *rif* transcription started as *var* transcription trailed off at about 12 hours post invasion (hpi) and peaking at 18hpi in the early pigmented trophozoite stage (Kyes et al., 2000). *Stevor* expression was shown to peak in trophozoites (Kaviratne et al., 2002). More recent studies involving whole transcriptome analysis by RNAseq (Otto et al., 2010) and microarrays (Le Roch et al., 2003, Rovira-Graells et al., 2012, Llinas et al., 2006) have detected *rif* and *stevor* transcripts in multiple IDC stages. Bachmann and group using quantitative real-time PCR reported that there were 2 peaks of the 2TM gene family expression in wild and laboratory isolates. In ex vivo parasites *var*, *rif*, *stevor* and *pfmc2tm* transcripts were detected in early rings. *Var* transcripts were maintained at high level until about 20hpi while 2TM transcription decreased to a minimum at 12-16hpi before rising in a second peak at the mid trophozoite stage (24-32hpi) (Bachmann et al., 2012). They observed the same 2TM gene variants in the two peaks of expression, the ring and trophozoite stages. In the same study wild isolates showed higher levels of VSA

expression than the laboratory isolate 3D7. This chapter will focus on *rif* and *stevor* transcript detection across IDC in laboratory and wild isolates by reverse transcriptase PCR (RT-PCR), and cloning and sequencing.

4.2. Objectives.

i.) To analyse *rif* and *stevor* expression profiles in wild and laboratory isolates in order to identify peak expression points as well as identify dominant transcripts.

ii.) Utilize Northern blotting analysis to determine whether transcripts detected by RT-PCR were properly spliced.

4.3. Method.

The methods in this chapter have been detailed in Chapter 2 (materials and methods). *Rif* and *stevor* transcription was analysed in *P. falciparum* laboratory and wild isolates (Table 4.1). Laboratory isolates (3D7 and IT) and culture-adapted Kisumu isolates, rosetting clone (SA075R+) and non-rosetting fraction (SA075R-) were thawed from frozen glycerolyte stocks and maintained in culture to bulk up IE. RNA was extracted from at least 100ul of RBC pellet per time point of synchronous cultures at 5% parasitemia. Double sorbitol synchronization was carried out before commencing the time course experiment. For 3D7 and IT, RNA was extracted at 6 time points, corresponding to early ring (ER), late ring (LR), early pigmented trophozoite (ET), mid

trophozoite (MT), late trophozoite (LT) and schizont (S) stages. For the Kisumu isolate SA075 RNA was extracted at 4 time points, ring (R), late ring to early pigmented trophozoite (LRET), mid to late trophozoite (T) and schizont (S) (Appendix table 9.5 and 9.6). For the Kilifi ex vivo isolates the time course experiment was carried out by culturing the ex vivo isolates for a single IDC. RNA was extracted at 7 time points (T1 to T7) every 10 hours across the IDC and certain time points were selected for expression profiling of *pir* genes (Appendix table 9.5). For some parasite isolates the sampling extended to the next IDC, with last time point (T7) coinciding with the second-generation rings. The time points were selected to represent the developmental stages; ring, late ring/ early trophozoite, mid/ late trophozoite and schizont stages, depending on availability of good quality RNA. Genomic DNA (gDNA) was extracted from asynchronous cultures at trophozoite stage.

Table 4.1. *Isolates used for rif and stevor expression profiling by cloning and capillary sequencing.*

Isolate ID	Isolates for *rif/stevor* expression during asexual blood stage growth	cDNA time points analysed	*rif* sequences	*stevor* sequences
3D7	Laboratory isolates (48hr cycle RNA time series)	6	✓	-
IT		4	✓	-
SA075c_R+	Culture-adapted Kisumu isolate; Rosetting positive (R+) or negative (R-)	4	✓	✓
SA075c_R-		4	✓	✓
10594#	Kilifi isolates from non-severe malaria (acute/*ex vivo* 48hr cycle RNA time series; M. Mackinnon/M. Rono collaboration)	3	✓	-
10595#		4	✓	-
10735		N/A	✓	-
10739		4	✓	✓
10761		2	✓	✓
10814		2	✓	✓

#: Isolates for which cDNA was prepared without DNase digestion to remove gDNA contamination. For complete parasite metadata in appendix table 9.6 and detailed parasite staging in appendix table 9.5.

RNA harvested at the different time points was treated with DNase to remove any gDNA contamination prior to cDNA synthesis. Two cDNA samples were prepared per RNA sample, one sample with (RT+) and one without (RT-) the reverse transcriptase. The RT- cDNA sample was set up to check for gDNA contamination in the RNA. For two Kilifi ex vivo isolates (10594 and 10595) already prepared cDNA without DNase treatment was obtained. Reverse transcription PCR (RT-PCR) was done using gene family specific primers (Figure 5.2 and Appendix Table 9.7). A complex set of degenerate primers, 3 forward and 1 reverse (Kyes et al., 1999), were used to sample *rif* repertoires and transcription, and a set of published forward and reverse primers were used for stevor (Lavazec et al., 2006). For each isolate the two cDNA amplicons, RT+ and RT- were run in parallel on an Agarose gel, in adjacent wells.

***Table 4.2**. rif and stevor primer sets. A complex set of 4 primers, 3 forward (rifF1, rifF2 and rifF3) and 1 reverse (rifR) were used to generate rif amplicons. Lavazec L and R primers were used to generate stevor amplicons.*

Primer	**Sequence**
rifF1 fwd	5' – C(AG)T CAC GA(GT) TGT TAA GCG
rifF2 fwd	5' – CGA (AG) (CT)G TGA ATT GTA TGC
rifF3 fwd	5' – C(CT)A C(CT)A G(AG)T TAT TAT GCG
rifR rev	5' – CTT CA (AT) ATT (AG)TT (AT)TT T(CT) (GT) (AGT)CG ATA ACG
Lavazec_L	5' - CAA AAG GAA GAG ATA AGT AT
Lavazec_R	5' - GTT TCT TGC ATT CAT GTT TCC

Prior to cloning *rif* and *stevor* amplicons were purified as described in Chapter 2. For *rif* two approaches were used for purification, total amplicon or gel extraction of specific bands. For *stevor* total amplicon purification using the QIAquick PCR Purification Kit was carried out. Cloning and screening of positive pcr2.1TOPO vectors into pcr2.1TOPO was carried out by restriction digestion using ECOR1 as described in Chapter 2.6.5 to 2.6.6.

Capillary Sanger sequencing of insert-positive vectors was done using M13 forward and reverse primers. Sequencing was carried out in-house on the ABI 3130 sequencer. Sequence assembly and analysis was carried our as described in Chapter 2. Briefly forward and reverse sequence reads were assembled into contigs using CAP3 in BioEdit and Geneious pro. Multiple alignments of the contigs were done on Geneious pro using MUSCLE. Vector and primer sequences were trimmed off. Contigs that met a 99% sequence identity threshold were clustered using CD-HIT. BLAST analysis was run on clusters representatives on Geneious pro. Proportions of clusters per sample were analysed for each isolate and expressed as pie charts.

4.4. Results

4.4.1. *Rif* and *stevor* transcripts detection by RT-PCR

Rif and *stevor* transcripts were sampled by reverse transcriptase PCR (RT-PCR) using sets of primers described in the methods and table above and detected by gel electrophoresis. *Rif* and *stevor* amplicons were detected at all time points across the IDC in all parasite isolates (examples Figure 4.1 and 4.2). This contrasts with previous work that suggested *rif* and *stevor* transcription occurred within a narrow window (Kaviratne et al., 2002, Kyes et al., 2000). Amplicons of the expected size ranges were observed, 800-980 base pairs (bp) for *rif* and 400bp for *stevor* (Figure 4.1 and 4.2). No amplicon was detected in the RT- sample confirming that there was no gDNA contamination in the cDNA prep.

A striking observation was the difference in band patterns and intensities for *rif* in different time points and in different isolates (Figure 4.1). IT cDNA amplicons appeared as a single band (Figure 4.1. A) while 3D7 had 2 distinct bands (Figure 4.1. B). Wild isolates had complex banding patterns with up to 3 distinct bands in some stages (Figure 4.1. C and D). The observation of multiple bands indicates transcription of multiple *rifs* at a single time point. The upper and lower bands most likely corresponded to transcripts from the 2 *rif* subgroups, longer *rifA* and shorter *rifB* (discussed in Chapter 3). This would suggest that in some isolates members of the two *rif* groups are co-expressed during the IDC. gDNA amplicons appeared as a smear on Agarose gel indicating that multiple *rif* variants were amplified by our primers.

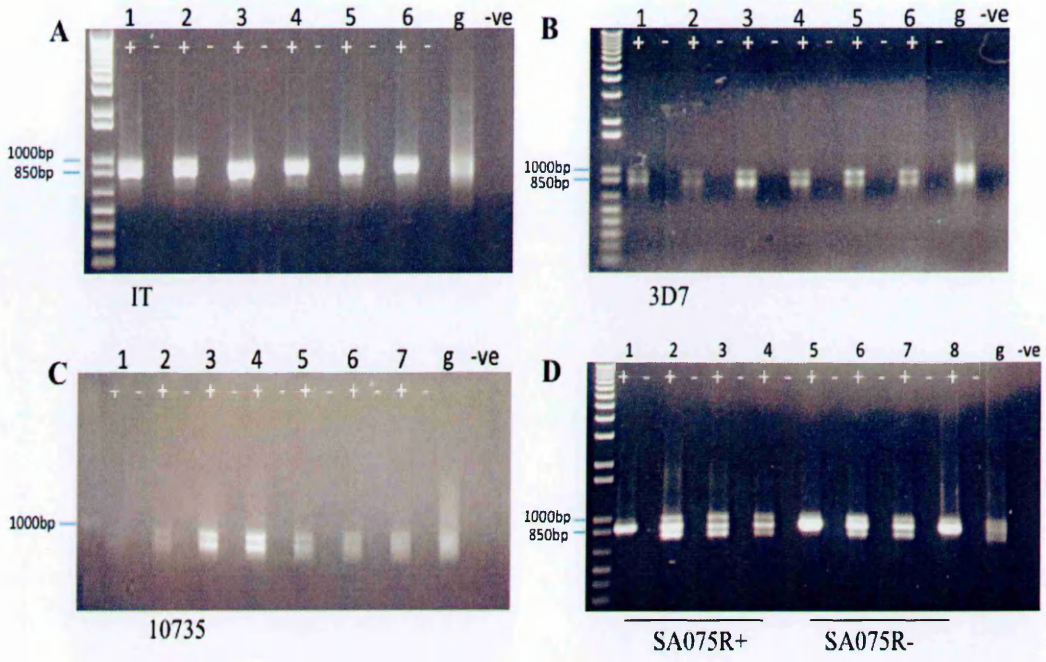


Figure 4.1: *rif amplicons detected on 1% agarose gels. Top panel: laboratory isolates IT and 3D7 (A and B) amplicons at 6 different time points and gDNA. Lower panel (C) wild isolate 10735 amplicons at 7 different time points and gDNA. (D) Culture-adapted wild isolates, rosetting clone SA075R+ and non-rosetting fraction SA075R- amplicons at 4 different time points and gDNA from SA075R+. (+) and (-) indicate samples with (RT+) and without (RT-) the reverse transcriptase enzyme. The product size in base pairs (bp) is indicated.*

Stevor amplicons of the expected (400bp) were observed across all timepoints. Unlike *rif* a single band was observed in all isolates at all time points and in gDNA (Figure 4.2) *stevor* sequence lenghts are under normal distribution therefore the amplicons are likely to be of similar lengths as observed on the gel images. Non-specific bands of about 200bp and 300bp were observed in cDNA in some isolates (Figure 4.2. B and C), while upper non-specific bands were detceted in gDNA (Figure 4.2).

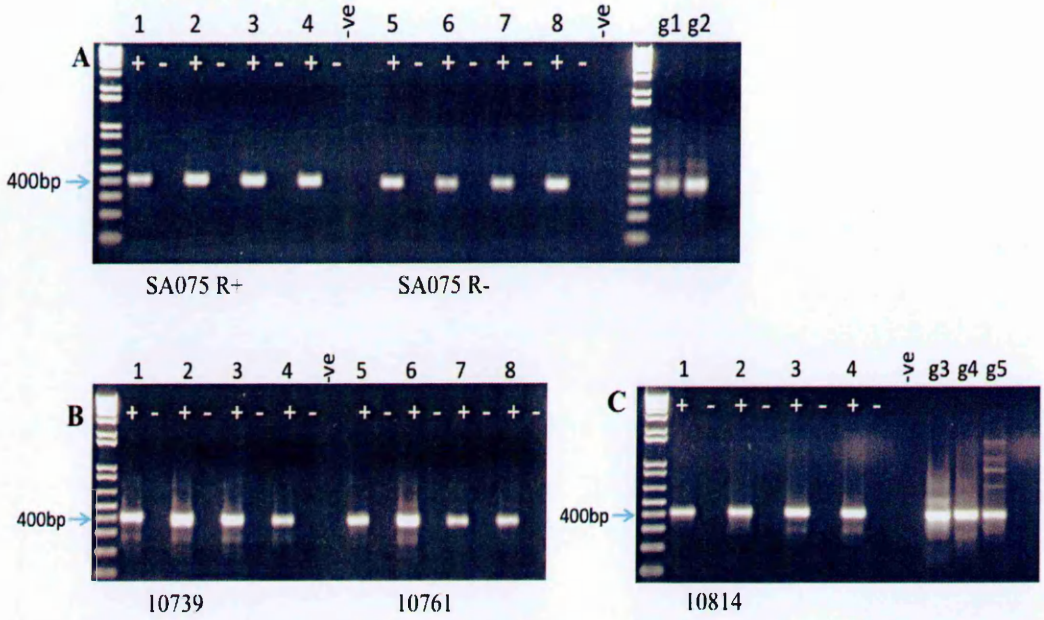


Figure 4.2. *1% agarose gel electrophoresis images of stevor amplicons in culture-adapted and wild isolates. (A) Amplicons from culture adapted isolate SA075; rosetting clone R+ (1-4) and non-rosetting uncloned fraction R- (5-8) cDNA and gDNA (g1 and g2 respectively). Lower panel: amplicons from wild isolates 10739, 10761 (B) and 10814 (C) cDNA at four time points across the IDC and gDNA (C) g3, g4 and g5. The '+' or '-' signs indicate samples with and without the reverse transcriptase enzyme respectively. The product size in base pairs (bp) is indicated.*

4.4.2. Cloning and sequencing *rif* and *stevor* amplicons.

Rif and *stevor* amplicons were cloned and sequenced as previously described in the methods chapter. 1595 *rif* and 498 *stevor* contigs were successfully assembled. These were clustered into 537 *rif* and 58 *stevor* clusters. Sequence searches using BLAST were carried out on cluster representatives.

4.4.2.1. *rif* cloning bias.

From an initial analysis of *rif* sequences obtained from cloning total PCR from a wild isolate a significant bias towards recovery of *rifB* sequences was observed (Figure 4.3). On average *rifA* are longer thar *rifB*, 1227 and 1115 basepairs (bp) for *rifA* and *rifB* respectively from 3D7 genome (see Chapter 3 section 3.3). It is therefore likely that the shorter *rifB* amplicons were preferentially cloned into the TOPO.

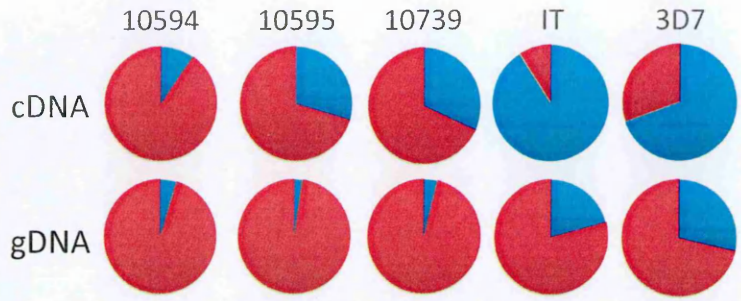


Figure 4.3. *Proportions of rifA and rifB recovered in cDNA (top panel) and gDNA (bottom panel) from cultured and wild isolates by rif amplicon cloning and sequencing. Significant bias towards recovering rifB sequences (red) in gDNA for all isolates and in cDNA of wild isolates was observed. rifA (blue) was predominantly recovered in cDNA from laboratory isolates.*

To further explore the extent of this bias comparison was made between cloning total amplicon and cloning gel extracted upper and lower bands separately (see Figure 4.1. B, C and D). Total amplicon as well as gel extracted upper and lower bands from 3D7 (Figure 4.1. B lane 5+) and 10735 (Figure 4.1. C lane 4+) trophozoites were cloned and sequenced. These two samples were selected on the basis that the upper and lower bands appeared to be of similar intensities (Figure 4.1. B and C). Length analysis gave some indication that there might be a bias was towards shorter sequences when total amplicon was cloned (Table 4.3). In 3D7 the average length of *rif* sequences from total amplicon cloning was 865bp as compared to 896bp in the upper band, and in the wild isolate 891bp and 909bp from total amplicon and upper band respectively. On average mostly *rifA* sequences were rcovered from cloning the upper band and total PCR product, while the lower band gave mostly *rifB* sequences (Table 4.3).

Table 4.3 *summary of the lenght and type of rif recovered from cloning total amplicon versus seperately cloning gel extracted upper and lower bands.*

	3D7 late trophozoites			10735 early trophozoites		
	Upper band	Lower band	Total	Upper band	Lower band	Total
Total *rif*	896bp (15)	823 (15)	865 (15)	909 (16)	830 (18)	891 (17)
rifA	912 (12)	814 (6)	865 (15)	924 (14)	820 (13)	897 (14)
rifB	832 (3)	830 (9)	0	833 (2)	850 (5)	961 (3)

Although the results above suggested cloning bias the analysis is based on sequences obtained from cDNA and may be confounded by expression bias. Analysis based on sequences from gDNA sequences was not possible as bands from gDNA did not separate well on the gel and appeared as a smear (Table 4.4).

To confirm that indeed shorter *rifs* were preferentially recovered from cloning total amplicon, a comparison of lengths was done between *rif* recovered by cloning total 3D7 gDNA amplicons and the expected length of *rif* amplicon in 3D7 genome. We observed a bias towards recovery of shorter *rif* when total amplicon was cloned than would be expected (Table 4.4). We observed no specific bias towards recovering A or B type *rif*. To overcome the length bias during cloning we gel purified the upper and lower bands separately in subsequent cloning and sequencing experiments.

Table 4.4. *Table comparing sequence lengths of total rif from 3D7 genome, expected rif amplicons based on primer coverage and observed coverage from 3D7 gDNA cloned amplicon.*

	No. of sequences	mean length	shortest	longest
Total *rif*	184	847.3	645	968
rif amplicon [#] expected	163	844.5	728	953
rif amplicon[@] observed	35	829.2	742	939

#: expected rif coverage by the complex set of rif primers obtained by running primer test (Primer3 application on Geneious Pro).

@: *rif* sequences obtained from cloning total 3D7 gDNA amplicon

4.4.3. Primer coverage from cloning and capillary sequencing data.

rif primer coverage was determined by computing the number of unique sequences obtained from cloning and sequencing *rif* amplicons from 3D7 gDNA. *Stevor* primer coverage was analysed from SA075 gDNA amplicons. The expected coverage was based on the 3D7 genome repertoire of 184 *rif* copies and 40 *stevor* copies. For the *rif* primers, the expected coverage was 17% due to the limited number of clones analysed (35 clones). To obtain 80% coverage analysis of at least 400 clones from gDNA per experiment was required, but this was not feasible at the time. Despite limited sampling the there was no significant difference between expected and observed coverage for *rifB* (Figure 4.4). There was a large difference between expected and observed *rifA* mainly because very few *rifA* were recovered from cloning and sequencing total gDNA amplicon (Figure 4.4). It was therefore difficult to establish primer bias for *rifA*. For *stevor* primers (Lavazec, 2006) the expected coverage in 61 clones sequenced (78%) was much greater than the observed, (38%) pointing to primer bias (Figure 4.4).

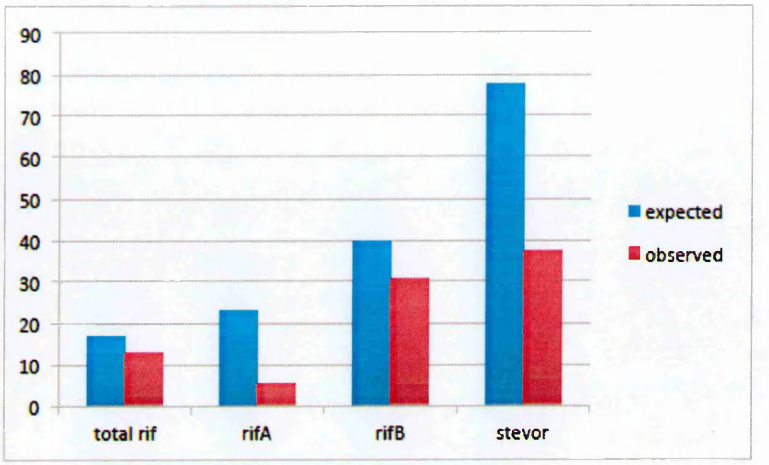


Figure 4.4. *Primer coverage for rif/stevor expression analysis by cloning and sequencing 3D7 gDNA (rif) and SA075 gDNA (stevor). Blue bars: expected coverage (%), red bars: observed coverage (%).*

4.4.4. *Rif* transcription profiling in laboratory isolates.

There was high diversity of *rif* transcripts expressed throughout the intra-erythrocytic developmental cycle (IDC) in both the field and laboratory isolates. In 3D7 there were different dominant transcripts at rings, trophozoites and schizonts (Figure 4.5 and Table 4.5). In rings the dominant *rif* PF3D7_1300400 (PF13_0004) belongs to the conserved UpsA *rif* group that are in a head-to-head orientation with *UpsA var* genes. There was a switch to different *rif* transcripts at the middle stages, with a single dominant *rif* PF3D7_0425900 (PFD1240w) at the early pigmented trophozoite stage, the stage of peak *rif* expression. At the schizont stage there was a switch to different *rif* PF3D7_1000300 (PF10_0003) and PF3D7_1373400 (MAL13P1.535) but the number of sequences recovered was too small to confidently identify dominant transcripts. Two minor transcripts were over represented in gDNA suggesting some level of bias to certain variants.

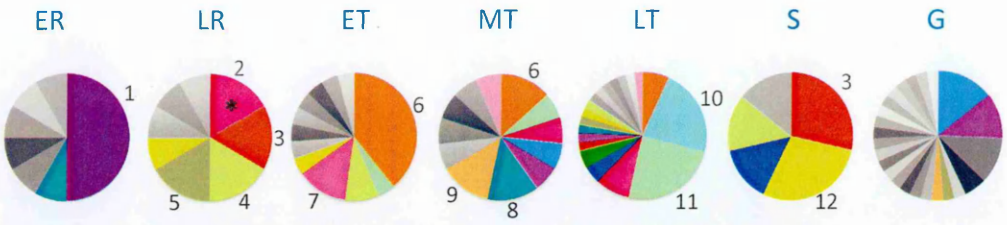


***Figure 4.5**. 3D7 rif genomic repertoire from 35 sequences (G), and transcription (cDNA) profile across IDC; early ring (ER), late ring (LT), early trophozoite (ET), mid trophozoite (MT), late trophozoite (LT) and schizont (S). Coloured regions of the cDNA pies represent sequences that were recovered more than once from cDNA while in the genomic pie coloured regions represent sequences that came up at least once in cDNA. Numbered regions represent the dominant transcripts in the specific stage (sequence labels in Table 4.5).*

***Table 4.5.** Dominant rif transcripts in 3D7 cDNA.*

Label	Cluster ID	Frequency % (proportion) at peak expression point	Peak expression point	A/B	BLAST top hit
1	RIF_K044	50% (6/12)	Early ring	A	PF13_0004
2	RIF_K038*	17% (2/12)	Late ring	A	PFD1020c|ps
3	RIF_K048	17% (2/12) and 29% (2/7)	Late ring and schizont	A	MAL13P1.535
4	RIF_K065	17% (2/12) and 9% (2/23)	Late ring and early trophozoite	A	PFA0760w
5	RIF_K084	17% (2/12)	Late ring	A	PF3D7_0701100
6	RIF_K015	39% (9/23)	Early trophozoite	A	PFD1240w
7	RIF_K099	13% (3/23)	Early trophozoite	A	PFI1825w
8	RIF_K079	13% (2/15)	Mid trophozoite	B	PF14_0003
9	RIF_K103	13% (2/15)	Mid trophozoite	B	PF14_0770
10	RIF_K022	23% (10/45)	Late trophozoite	B	PF13_0006
11	RIF_K034	25% (11/45)	Late trophozoite	A	PFL0025c
12	RIF_K151	29% (2/7)	Schizont	B	PF10_0003

*/ |ps: *pseudogene.*

Rif transcription profiles were analysed for four time points across the IDC in IT and at an additional 2 time points in a biological replicate that was passaged in culture for over 40 cycles (ITp). IT showed much more focused *rif* expression profile with a single dominant transcript at the later stages (Figure 4.6. A). PFIT_bin02300 (1), PFIT_0835500 (2) and PFIT_0424300 (3) were the dominant ring transcripts with PFIT_0835500 dominating trophozoite and schizont stages. IT *rif* PFIT_0835500, PFIT_bin02300 and PFIT_0734200 are 99% identical to *rif* that were previously identified in Palo Alto by Kyes and group (Kyes et al., 1999). PFIT_0835500 is identical to *rif1* that was expressed in Palo Alto clones in both the rosetting and non-rosetting fractions, PAR+ and PAR- respectively while PFIT_0734200 and PFIT_bin02300 are identical to *rif2* and *rif3* respectively expressed in the non-rosetting fraction PAR-. None of the transcribed *rif* were recovered from gDNA (Figure 4.6. A. G) but this could be due to under sampling the genomic repertoire (only 24 sequences were recovered from cloning and sequencing IT gDNA amplicon). A different set of *rif* emerged in IT after continuous passage in culture (Figure 4.6. B). This could have been caused by slight modification in the reverse primer used to amplify *rif* from the passaged IT (see discussion below). The reverse primer was modified to make it more degenerate, and therefore able to capture more *rifs*. A pseudogenic transcript PFIT_0536700 (5) was predominantly recovered in ITp rings and trophozoites. This pseudogene is closely related to a conserved 3D7 pseudogene (PF3D7_0533000/PFE1635w) that lies tail-to-head with the *var1csa* pseudogene (PF3D7_0533100) on the chromosome. The *rif* PFIT_0424300 (3) was recovered in rings and trophozoites of both IT and ITp while PFIT_0900150 (4) was mainly found in trophozoites.

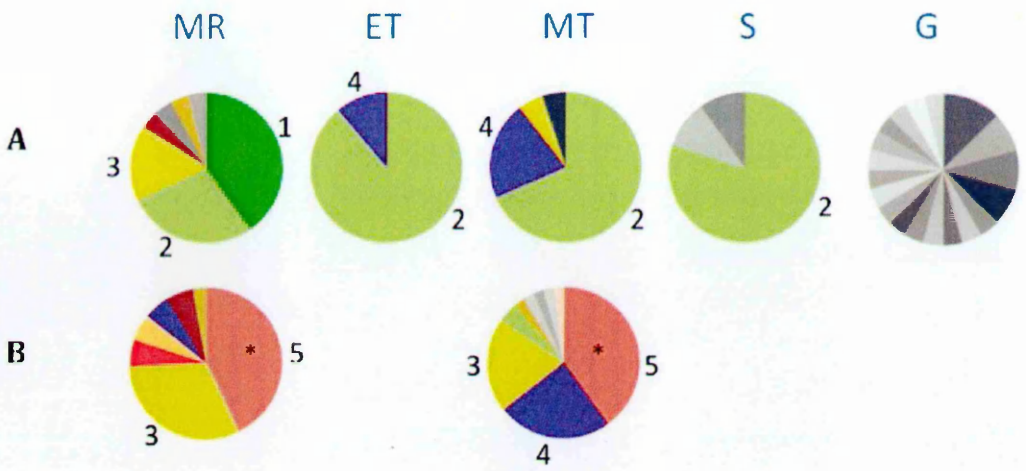


***Figure 4.6**. rif genomic repertoire from 24 sequences (G) and transcription (cDNA) profile in (A) Original IT line at 4 IDC stages; mid ring (MR), early trophozoite (ET), mid trophozoite (MT) and schizont (S), and in (B) ITp passaged continuously in culture for 2 months. Coloured regions of the cDNA pies represent sequences that were recovered more than once from cDNA while in the genomic pie coloured regions represent sequences that came up at least once in cDNA. Numbered regions represent the dominant transcripts in the specific stage (sequence labels in Table 4.6).*

***Table 4.6.** Dominant rif transcripts in IT cDNA.*

Label	Cluster ID	Frequency % (proportion) at peak expression point	Peak expression point	A/B	BLAST top hit
1	RIF_K021	40% (10/25)	Ring$^{\alpha}$	A	PFIT_bin02300
2	RIF_K004	89% (8/9), 68% (13/19) and 80% (8/10)	Late ring to Schizont$^{\alpha}$	A	PFIT_0835500
3	RIF_K050	31% (11/35)	Ring$^{\beta}$	A	PFIT_0424300
4	RIF_K055	21% (4/19) and 24% (11/45)	Trophozoites $^{\alpha}$ and $^{\beta}$	A	PFIT_0900150
5	RIF_K532	43% (15/35) and 40% (18/45)	Ring$^{\beta}$ and trophozoite $^{\beta}$	A	PFIT_0536700*

α: Original IT

β: ITp after 2 months of continuous passage in culture.

**: pseudogene*

4.4.5. *Rif* transcription profiling in wild isolates.

There was high diversity of *rif* transcripts in wild isolates across the IDC with little sharing between isolates (4.7). It was not possible to determine dominant transcripts in some time points. A few conserved transcripts were commonly expressed in wild isolates. A conserved *rifB* RIF_K006 (1) was expressed in 4 of the 6 Kilifi isolates and recovered from gDNA in the Kisumu isolate. This gene is 96% identical to 3D7 PFE1630w, one of the conserved *rif* with orthologs identified in HB3, DD2 and a Brazilian field isolate (Claessens et al., 2011). Another *rifB* RIF_K037 was expressed in 2 Kilifi isolates at the trophozoite stage and found in the genomic DNA of the same 2 Kilifi isolates and in 3D7. This *rif* is related to 3D7 *rif* PFF1570w/PF3D7_0632300 (94% identity). A conserved *rifB* pseudogene RIF_K014 (2) was recovered from 10594 rings and schizonts and 10814 gDNA. This pseudogene is highly conserved with 100% identity to a 3D7 pseudogene PFE0015c/PF3D7_0500300. A late stage *rifB* RIF_K025 (4) dominated expression in 10595 schizonts and 10793 trophozoites. This *rif* is closely related to a Brazilian *rif* DQ265287 and a 3D7 *rif* PFD0045c.

In one isolate 10761 a single *rifA* RIF_K011 (10) dominated ring and schizont transcription. This *rif,* PFL2585c is strain transcendent (Claessens et al., 2011). In a previous study PFL2585c was upregulated in clinical isolates as compared to laboratory isolates (Lemieux et al., 2009). In this isolate different sets of transcripts were expressed in the mid stages of the IDC (Figure 4.7 panel 5). In the isolate 10814 a clearer pattern of stage-specific *rif* transcription emerged. This was probably because of better sampling depth. Different *rif* transcripts were upregulated in the different IDC stages (Figure 4.7

panel 6). Two *rifA* dominated at the ring stage, a different *rifA* came up in the ring to trophozoite transition stage, a *rifB* dominated in trophozoites and a second *rifB* in schizonts.

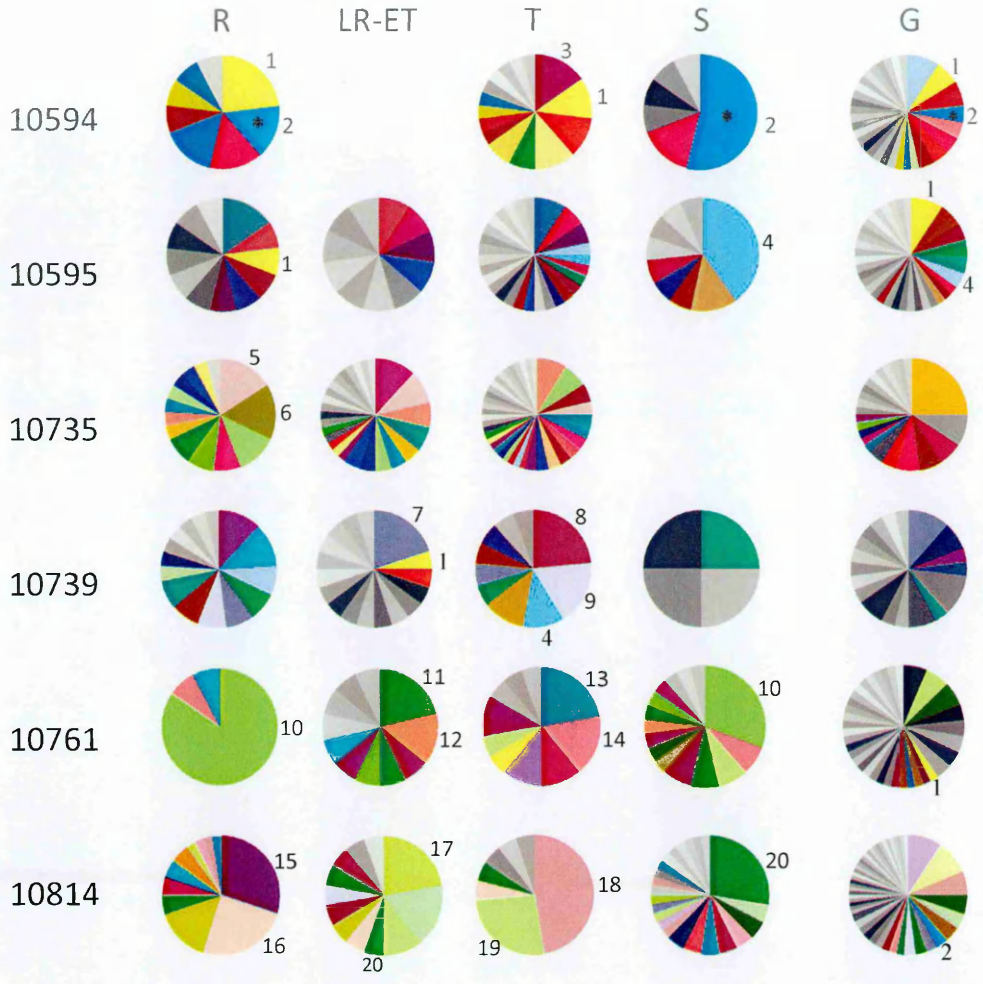


Figure 4.7*. rif genomic repertoire (G) from 43, 44, 44, 26, 47 and 96 sequences from 10594, 10595, 10735, 10739, 10761 and 10814 respectively, and transcription (cDNA) profile at 4 IDC stages; ring (R), late ring- early trophozoite (LR-ET), trophozoite (T) and schizont (S), in wild isolates. The coloured regions represent sequences that were recovered more than once in cDNA. Numbered regions represent the dominant transcripts and the time points or gDNA from which they were recovered. (Sequence labels and frequencies in Table 4.7).*

Table 4.7. Dominant rif transcripts in wild isolates.

Label	Cluster ID	Frequency % (proportion) at peak expression point	Peak expression time point	A/B	BLAST top hit
1	RIF_K006	23% (3/13)	Ring[1]	B	PFE1630w
2	RIF_K014	54% (7/13)	Schizont[1]	Bps	PFE0015c|ps
3	RIF_K089	15% (4/26)	Trophozoite[1]	B	PFB1005w
4	RIF_K025	40% (6/15)	Schizont[2]	B	PFD0045c
5	RIF_K019	16% (4/25)	Ring[3]	B	PFF1570w
6	RIF_K049	16% (4/25)	Ring[3]	A	PFI0030c[b]
7	RIF_K030	20% (4/16)	Ring[4]	B	PFC0030c
8	RIF_K070	24% (4/17)	Trophozoite[4]	A	MAL13P1.535[b]
9	RIF_K057	18% (3/17)	Trophozoite[4]	A	PFB0015c
10	RIF_K011	85% (11/13) 31% and (8/26)	Ring5 and schizont[5]	A	PFL2585c
11	RIF_K069	21% (3/14)	Late ring/ early trophozoite[5]	A	PFD1240w[b]
12	RIF_K126	14% (2/13)	Late ring/ early trophozoite[5]	A	PFI0010c[b]
13	RIF_K043	22% (4/17)	Trophozoite[5]	B	PF10_0397
14	RIF_K032	18% (3/17)	Trophozoite[5]	B	PFF0045c|ps
15	RIF_K016	30% (12/40)	Ring[6]	A	PF11_0021[b]
16	RIF_K018	25% (10/40)	Ring[6]	A	PFL2625w[b]
17	RIF_K058	22% (4/18)	Late ring/ early trophozoite[6]	A	PFB1035w
18	RIF_K009	47% (9/19)	Trophozoite[6]	B	PFB0055c[b]
19	RIF_K027	26% (5/19)	Trophozoite[6]	A	PFA0760w[b]
20	RIF_K013	28% (11/40)	Schizont[6]	B	PFC0030c[b]

[1]: 10594, [2]: 10595, [3]: 10735, [4]: 10739, [5]: 10761, [6]: 10814

[b]: Top BLAST with low sequence similarity.

4.4.6. *Rif* transcription profiling in a rosetting laboratory-adapted Kisumu isolate.

Results from *rif* and *stevor* expression analysis in the rosetting wild isolate SA075 revealed differential expression of *rif* and *stevor* in the rosetting clone and the non-rosetting fraction. At ring and schizont stage both parasites shared some *rif* and *stevor* transcripts but at late ring and trophozoite stages we observed different dominant transcripts in the rosetting and non-rosetting parasites (Figure 4.8).

A single *rifA* RIF_K001 (1) clearly dominated ring and schizont transcription in both R+ and R- lines (Figure 4.8. A). In the R+ clone two dominant transcripts emerged in the trophozoites a *rifA* RIF_K003 (3) and a *rifB* RIF_K005 (4). The *rifA* has no closely related ortholog but shares a small region of homology with MAL7P1.217 and a Brazilian *rif* DQ265279. The dominant trophozoite *rifB* is a conserved *rif* with orthologs in 3D7 (92% in 758/818bp), HB3 (93% in 762/818bp) and a Kilifi isolate 10595 (99% in 814/821bp). In our 3D7 line this *rif* was expressed predominantly in late trophozoites (Figure 4.5 and Table 4.5).

A *stevor* pseudogene STE_K002 (1) was predominantly transcribed in rings and schizonts in both rosetting and non-rosetting parasites (Figure 4.8). This *stevor* was 98% identical to 3D7 stevor pseudogene PF3D7_0102100 In the R+ trophozoites a major *stevor* transcript STE_K003 (2) was detected. This stevor has closely related orthologs in 3D7 PFF1550w and IT PFIT_0100600 (97% pairwise identity over 356/367bp). A pseudogene STE_K005 (3) was predominantly transcribed in the late ring/ early trophozoites and trophozoites. This pseudogene has no orthologs but is closely related to

a *stevor* from a Kilifi isolate. A transcript STE_K005 (4) orthologous to PFL2635w a 3D7 *stevor* was detected at high levels in the late ring/ early trophozoite stage.

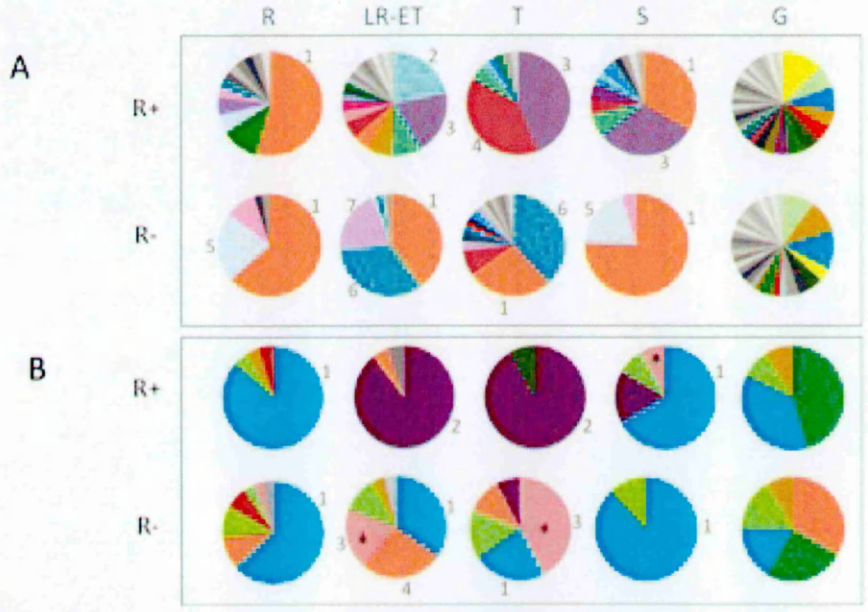


Figure 4.8. *pie charts showing genomic repertoires (G) of rif (A) and stevor (B) and expression profiles in a culture-adapted highly rosetting wild isolate across the IDC (R: rings, LR/ET: Late ring to early trophozoites, T: mid to late trophozoites and S: schizonts) in (R+) SA075 rosetting clone and (R-) non-rosetting fraction. For rif the genomic repertoires are based on 56 sequences (R+) and 50 sequences (R-) while for stevor 11 sequences (R+) and 12 sequences (R-). *: pseudogene.*

Table 4.8. *Dominant rif and stevor transcripts in a culture-adapted wild SA075.*

Gene	Label	Cluster ID	Frequency % (proportion) at peak expression point	Peak expression point	A/B	BLAST top hit
rif	1	RIF_K001	54% (19/35), 33% (14/43), 63% (26/41), 76% (31/42)	R+ and R- Ring and schizont	A	MAL13P1.500
rif	2	RIF_K012	22% (12/55)	R+ late ring/early trophozoite	A	PF3D7_0700200
rif	3	RIF_K003	44% (18/41)	R+ trophozoite	A	MAL7P1.217[b]
rif	4	RIF_K005	39% (16/41)	R+ trophozoite	B	PF10_0006
rif	5	RIF_K007	22% (9/42) and 20% (8/42)	R- ring and schizont	A	PF08_0138
rif	6	RIF_K002	34% (22/66) and 38% (24/42)	R- Late ring to trophozoite	B	PFB1005w
rif	7	RIF_K010	20% (13/55)	R- late ring/early trophozoite	A	RIF_K123
stevor	1	STE_K002	86% (19/22), 67% (8/12), 71% (12/17) and 88% (15/17)	R+ and R- ring and schizont	na	PFA0105w
stevor	2	STE_K003	89% (17/19) and 92% (23/25)	R+ Late ring to trophozoites	na	PFF1550w
stevor	3	STE_K006	42% (11/25)	R- Trophozoite	na	PF8443_8_3_U446600[b]
stevor	4	STE_K005	27% (7/19)	R- Late ring/early trophozoites	na	PFL2635w

[b] *: Top BLAST with low sequence similarity, 306/342 identical sites with 89% pairwise identity for the rif gene RIF_K003, and 302/380 identical sites with 79% pairwise identity for the stevor gene STE_K006.*

4.4.7. *Stevor* transcription in wild isolates.

Analysis of stevor sequences from wild isolates revealed a conserved and commonly expressed stevor pseudogene STE_K001 (1) that was recovered from all 3 Kilifi isolates at high frequency (Figure 4.9 and Table 4.9). The pseudogene was predominantly expressed in ring and schizont stages but was detected in late rings/ early trophozoites and in mature trophozoites. This stevor was 99% identical to a 3D7 stevor pseudogene PFA0105w (PF3D7_0102100); with two inframe stop codons occurring at the same location in the sequences from wild isolates and 3D7.

A second dominantly expressed stevor pseudogene STE_K015 (3) was detected at the ring to trophozoite transition stage in 10739. This stevor is closely related to an IT pseudogene PFIT_0100600 (98.5% identity across 837/850 nucleotides).

Other minor transcripts were detected at low frequency across the different isolates at different IDC stages. There was a significant primer bias for stevor, which can be seen in the gDNA pies (Figure 4.9. G).

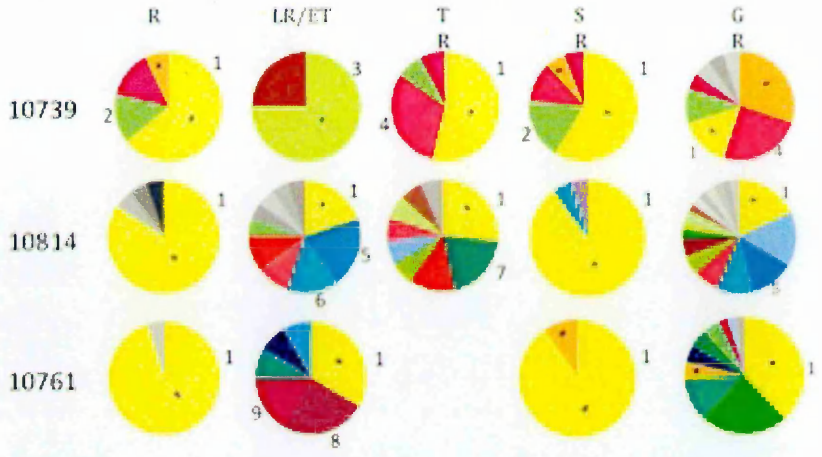


***Figure 4.9**. stevor genomic repertoire (G) from 20, 42 and 40 stevor sequences from 10739, 10761 and 10814 respectively, and, transcription (cDNA) profiles at 4 IDC stages; ring (R), late ring- early trophozoite (LR-ET), trophozoite (T) and schizont (S), in 10739 and 10814, and 3 stages in 10761. *: Pseudogene.*

Table 4.9. ***Dominant stevor transcripts in wild Kilifi isolates.***

Label	Cluster ID	Frequency %(proportion) at peak expression point	Peak expression point	BLAST top hit
1	STE_K001	64% (9/14) to 90% (18/20)	Ring and schizont1,2,3	PF3D7_0102100|ps
2	STE_K004	14% (2/14) and 18% (3/17)	Ring and schizont1	PF3D7_0400800
3	STE_K015	75% (6/8)	Late ring/ early trophozoite1	PFIT_0100600|ps
4	STE_K009	31% (4/13)	Trophozoite1	PFIT_bin00600
5	STE_K012	20% (4/21)	Late ring/ early trophozoite2	PFD0035c
6	STE_K014	15% (4/21)	Late ring/ early trophozoite2	PFIT_1100300
7	STE_K021	20% (3/15)	Trophozoite2	PFIT_bin05300[b]
8	STE_K023	25% (3/12)	Late ring/ early trophozoite3	PFIT_0800600
9	STE_K013	17% (2/12)	Late ring/ early trophozoite3	PF3D7_1254300

[b]: Top BLAST with low sequence similarity (192/216 (88%)

ps: pseudogene

4.4.8. Diversity of *rif* transcripts obtained from capillary sequencing.

The Simpson diversity index is a measure of homogeneity where a high SDI corresponds to more homogenous transcription. An analysis of the diversity of *rif* transcripts was carried out to determine whether the laboratory parasite isolates exhibited decreased transcription diversity than wild isolates due to the absence of immune selection pressure in culture. There was no significant difference in *rif* transcription diversity between laboratory and wild isolates, but this could be due to our low sampling depths as well as primer and cloning biases. IT and SA075 non-rosetting fraction showed the least diversity (0.284 and 0.289 respectively) while field isolates 10595 and 10735 showed the most diversity (0.030 and 0.031 respectively) with 3D7 showing moderate diversity (Figure 4.10). Interestingly rosette selected isolates IT, SA075 R- and SA075 R+ expressed markedly less diverse sets of *rif* transcripts. The IT clone used in our study obtained from the Sanger Institute had been put through rosette selection. Similarly the SA075 R+ was constantly kept under rosette selection. The R- fraction was however not put through negative rosette selection process, suggesting that the observed reduction in diversity was due to selection of a sub population of parasites that predominantly expressed a few subsets of *rif* genes. Overall most isolates showed higher diversity in *rif* recovered from gDNA as compared to cDNA, indicating overall good primer coverage. In one isolate 10735 there was higher transcript diversity than the recovered gDNA repertoire (Figure 4.10). This was probably due to under-sampling of the genomic repertoire (44 clones analysed for gDNA amplicon vs. 123 clones analysed for cDNA).

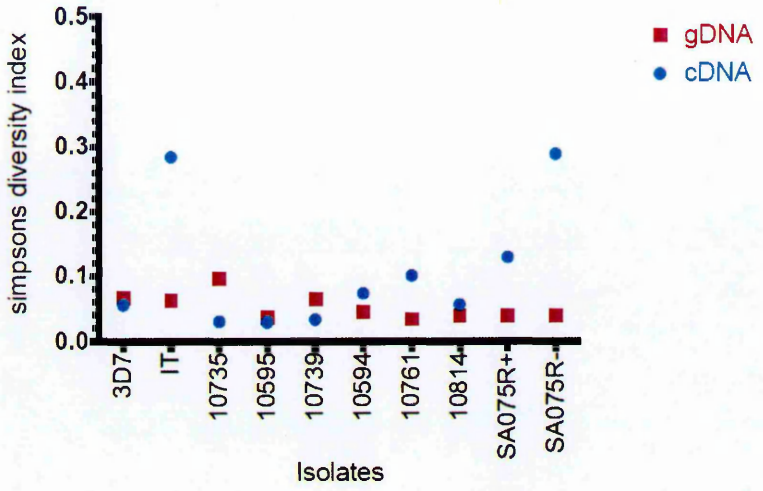


Figure 4.10. *rif transcript and genomic diversity from cloning and sequencing amplicons from cDNA (blue circles) and gDNA (red squares).*

4.4.9. Determining clone composition of Kilifi isolates using merozoites surface antigens 1 and 2 (MSP1 and MSP2).

Many isolates sampled from *P. falciparum* natural infections are composed of multiple distinct parasite clones. This multiplicity of clones in wild isolates could in part explain the higher diversity of transcripts observed in wild isolates as compared to laboratory isolates. To determine clonal composition of isolates MSP1 and MSP2 genotyping by capillary electrophoresis was carried out as described earlier (Liljander, 2009). As expected, the wild isolates were composed of multiple genotypes while the laboratory isolates had single clones each (Table 4.10). The number of clones in an isolate however did not completely explain the difference in the diversity of *rif* transcripts observed. (In this case high transcript diversity was defined by low SDI) (Table 4.10 and Figure 4.11). For instance two wild isolates 10595 and 10735 with different clonal diversity (4 and 2 respectively) showed similar SDI, 0.030 and 0.031 respectively. The laboratory line

composed of single clones showed moderate (3D7) and low heterogeneity (IT) in *rif* transcription (SDI of 0.56). This indicated that other factors besides clonal diversity may play a role in regulating *rif* transcription diversity.

Table 4.10. *Number of clones in isolates by MSP2 genotyping using capillary electrophoresis*

Label	Isolate	No. of clones by MSP2 genotyping			Diversity (SDI)
		FC27	IC	Total	
1	3D7	0	1	1	0.056
2	IT	0	1	1	0.285
3	SA075c_R+	0	2	2	0.130
4	SA075c_R-	0	2	2	0.289
5	10594	1	2	3	0.074
6	10595	3	1	4	0.030
7	10735	0	2	2	0.031
8	10739	4	0	4	0.034
9	10761	1	1	2	0.102
10	10814	1	1	2	0.057

SDI: Simpson's diversity index.

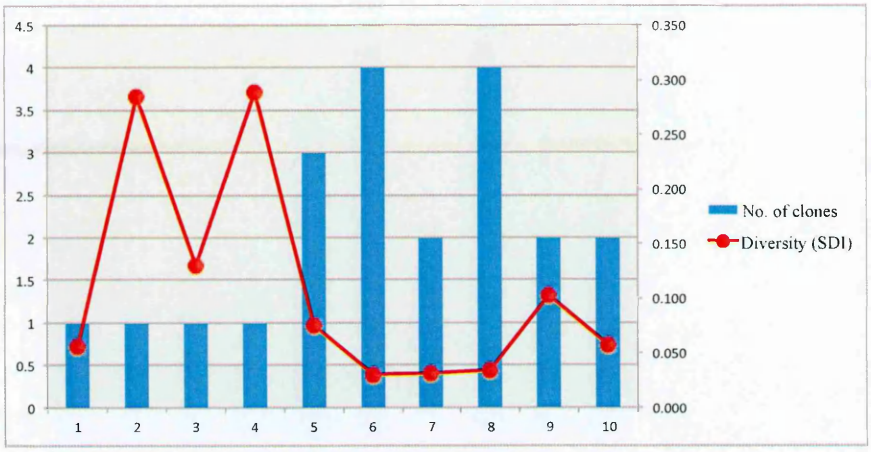


Figure 4.11*. Comparing number of clones and rif transcript diversity across isolates 1: 3D7, 2: IT, 3:SA075R+, 4:SA075R-, 5:10594, 6:10595, 7:10735, 8:10739, 9:10761 and 10:10814. Bar graph shows the number of clones by MSP1 genotyping (axis on the left) while the line graph shows the SDI of each isolate (axis on the right). High SDI corresponds to high homogeneity in transcription.*

4.5. Analysis of *rif* transcription by northern blotting.

The excessive diversity of *rif* transcript detected by the highly sensitive RT-PCR raised the question as to whether all the transcripts detected were full length and correctly spliced, between 1.8-2.1 kb range (Kyes et al., 2000). For this reason Northern blotting of RNA from laboratory isolates was carried out. A non-radioactive technique based on digoxygenin (DIG) labelling was applied because of the regulation against the use of radioactive isotopes in our laboratory.

RNA probes were gerenated by in vitro transcription using T7 RNA polymerase. RNA probes were preferred over DNA probes as they are better in detecting rare mRNA species in total RNA. RNA probes were generated by amplifying *rif*, *MSP1*, *varC* (var exon 2) and *calmodulin* genes using gene specific primers (Kyes et al., 2000), modified to include a T7 RNA polymerase sequence for in vitro transcription (Methods chapter section 2.10). *MSP1* and *var* probes were included to analyse stage-specific gene expression while calmodulin, a constitutively expressed gene was include to control for RNA amounts across the different parasites and IDC stages. *In vitro* transcription by T7 RNA polymerase was carried out using the DIG RNA Labelling Kit (SP6/T7) (Roche). Two sets of probes were generated, sense probes to detect antisense mRNA and antisense probes to detect sense mRNA, althoguh for the experiment only the anitsense probes were used. For sense probes the T7 RNA polymerase sequence was added to the froward primers while for the antisense probes the polyemares sequence was added to the reverse primers. Amplicons for probe transcription were generated from IT gDNA (Figure 4.12. A). Additional probes for *rif* and *calmodulin* were generated from cDNA from IT and

Kilifi isolates. For *rif,* cDNA amplicons were used to improve sesnitivity as only a small subset of *rif* are transcribed from the genomic repertiore. For *calmodulin* the transcript was much shorter than the genomic sequence (450bp and 982bp respectively) and therefore the shorter cDNA amplicon would be better labelled than the longer gDNA amplicon (Figure 4.12. A). Good labelling efficiency was obtained with intensities similar to those of the DIG-labelled RNA control included in the labelling kit (Figure 4.12. B).

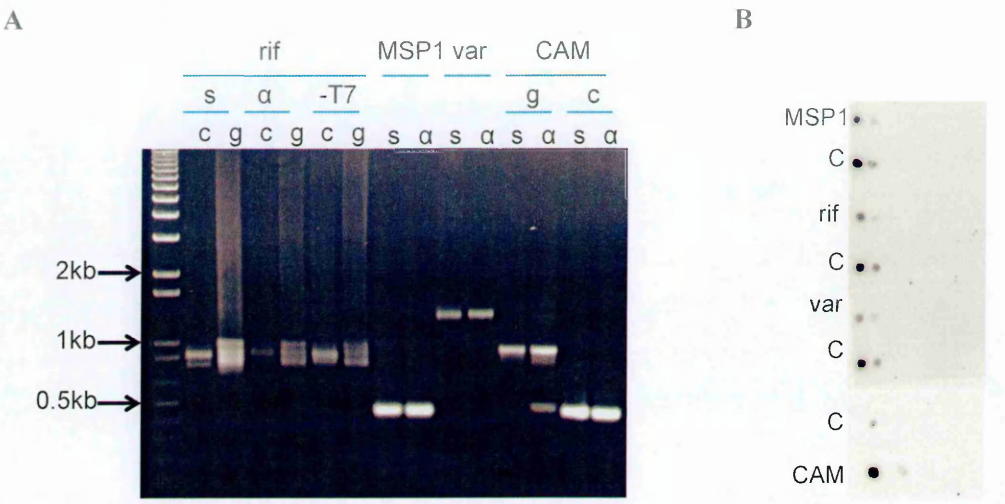


***Figure 4.12**: A: PCR products for probe generation. Rif, MSP1, var and calmodulin amplicons generated with primers containing T7 RNA polymerase sequence on the forward primer for sense (s) probes or the reverse primer for the antisense (α) probes. For rif and calmodulin (CAM) probes were generated from gDNA (g) and cDNA (c) amplicons. For rif a control PCR without the T7 promoter sequence in the primers (-T7) was carried out. B: Determination of labelling efficiency. Spot blots of dilutions of DIG-labeled probes, MSP1, rif (from gDNA), var and calmodulin (from gDNA) and the DIG-labeled RNA control. Only the first two dilutions of probes and control were detectable after 1200 seconds of exposure.*

1ug of total RNA was resolved on a 0.8% agarose gel (Figure 4.13. A) and successfully transferred to a nylon membrane, as shown by the DIG-labelled RNA marker (Figure 4.13 B and C). The 2.1kb 18S and 4.1kb 2.8S rRNA were clearly visible in the agarose gel (Figure 4.13. A) implying that RNA was intact with little degradation detected. When the DIG-labelled complex *rif* probe from IT gDNA was hybridized to the blot no signal was detected even after prolonged exposure upto one hour. A low intensity signal of the 18S and 28S rRNA was detected with the *rif* probe (Figure 4.13. B). No signal was detected when the blot was stripped and reprobed with the *MSP1* (Figure 4.13. C). In addition the signal from the RNA ladder was much weaker suggesting that signal intensity was significantly reduced by stripping (Figure 4.13. C).

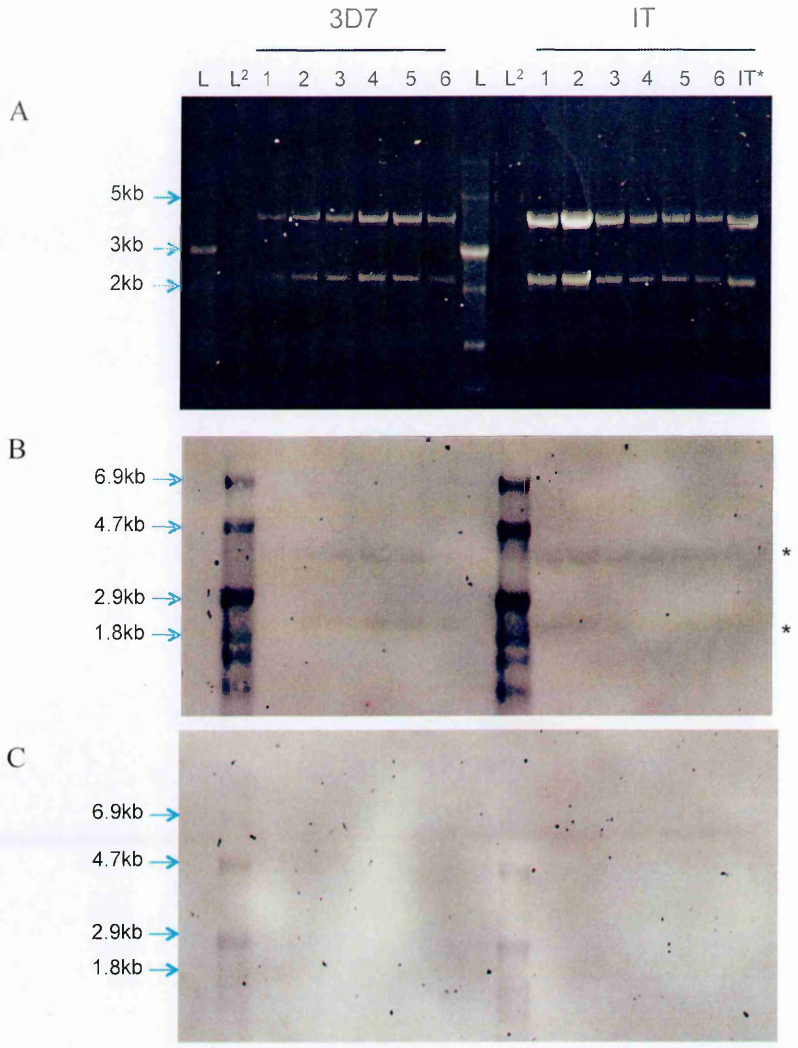


Figure 4.13: *Images of RNA 1ug of 3D7 and IT RNA from 6 time points of the IDC and an additional pooled RNA sample from IT (IT*) on a 0.8% agarose gel (A) and northern blots of the same RNA transferred onto a nylon membrane and probed with DIG-labbeled rif (B) and MSP1 (C) probes. L: single stranded RNA ladder (NEB). L^1: DIG-labbeled RNA marker I (Roche). *: non-specific 18s and 28s rRNA signale detected with rif probe.*

Failure to detect *rif* or MSP1 transcripts in the initial blot could have been as a results of low mRNA levels in the total RNA. To remedy this, two concentrations of total RNA from IT, a high concentration (5ug) and a low concentration (1ug) were assayed. The two high and low concentration RNA samples were loaded in quadruplicates together with the RNA ladders (Figure 4.14. A) to enable parrallel probing with the four different probes, complex *rif*, *var*, MSP1 and calmodulin without the need for stripping and reprobing. In addition, the parallel blots would reveal whether all probes gave the 18S and 28S non-specific rRNA signal.

A 6kb MSP1 mRNA signal was clearly detected in the 5ug RNA sample even after 5 minutes exposure. After 20 minutes exposure the signal from the 1ug RNA sample was visible (Figure 4.14. B MSP1 lane). A 1.3kb calmodulin mRNA band was decteted after 1 hour long exposure (Figure 4.14. C) thus the resulting image had excessive background signal. No signal was detected with *rif* and *var* probes even after prolonged exposure (Figure 4.14. B). The probes did label the amplicons that were used to generate them except the *var* PCR product (Figure 4.14. B). All probes labelled the 2.1kb and 4.1kb non-specific rRNA even in the low concentration (1ug) RNA sample suggesting that the signal was not probe specific.

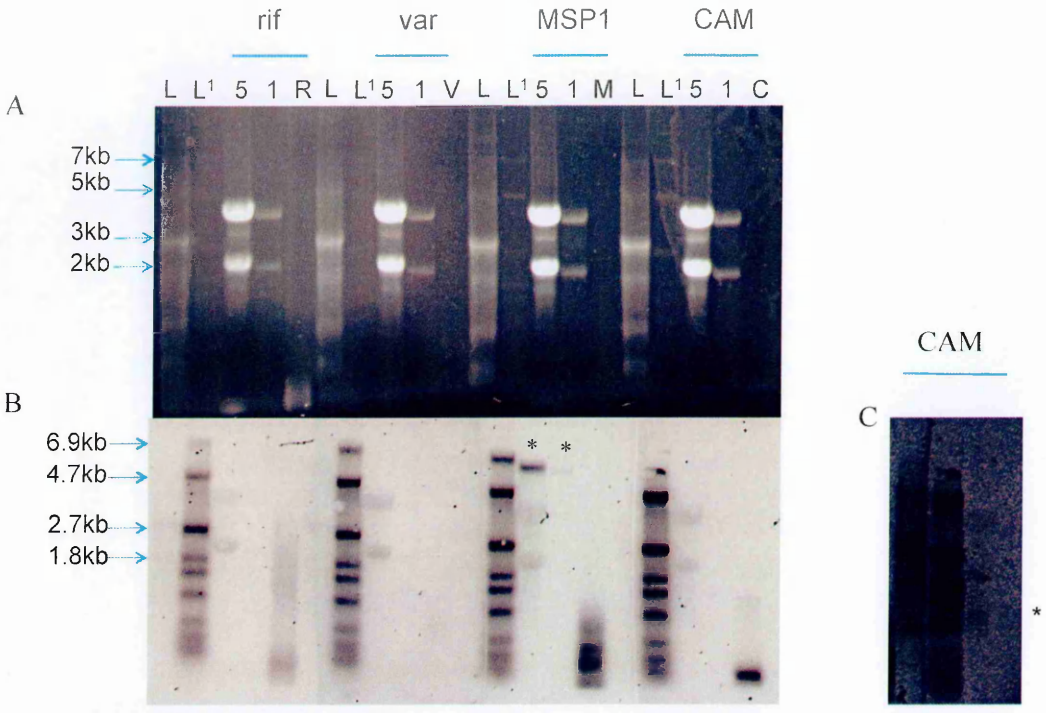


Figure 4.14: *Northern blotting of total IT RNA from 6 pooled timepoints of the IDC. (A) 0.8% ethidium bromide stained agarose gel with 5ug (5) and 1ug (1) of RNA, and 2ul of PCR product used to generate the four probes, rif (R), var (V), MSP1 (M) and calmodulin (C). (B) Blot of the RNA transferred onto a nylon membrane short exposure (20 minutes). The membrane was cut into 4 strips on the single stranded RNA ladder (L) lane and each strip probed with one of the four DIG-labelled probes. The asterisks (*) in figure B point to the MSP1 mRNA signal detected using the MSP1 probe in the high and low RNA concentrations. (C) A strip of the blot probed with calmodulin probe after long exposure (1 hour). The asterisks (*) in figure C point to a diffused signal from calmodulin mRNA. L: single stranded RNA ladder (NEB). L^1: DIG-labbeled RNA marker I (Roche).*

To rule out that the 18s and 28s rRNA band was as a result of non-specific binding of anti-DIG AP conjugate antibody to rRNA, a fresh blot was incubated with antibody before probibg with any of the four probes. No rRNA signal was detected in the 6ug, 4ug and 2ug of total RNA lanes (Figure 4.15). When the MSP1 probe was hybridised transcript was detected even in the 2ug RNA sample (Figure 4.15). No signal was detecetd with the *rif* and *var* probes even at high total RNA concentrations (Figure 4.14). Increasing probe concentraion, varying the amount of probe and antiDIG antobody, and exposure time had no effect on rif and var blotting results.

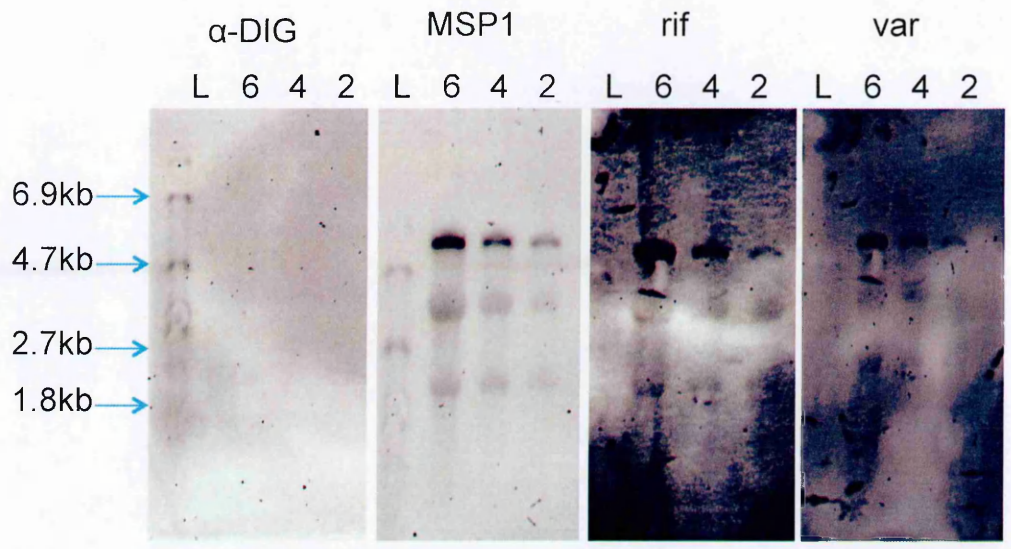


Figure 4.15: *Northern blotting of total RNA from pooled IT timepoints at 3 different concentrations 6ug (6), 4ug (4) and 2ug (2). In the first lane no probe was hybridiese to check specificity of the anti-DIG antibody. The second, third and fourth lanes is the blot sequentially probed with MSP1, rif and var respectively without stripping. L: DIG-labbeled RNA marker I (Roche).*

The *rif* multigene family is highly diverse and probes generated from gDNA amplicons might not be specific enough to detect transcripts. To make the assay more specific *rif* complex probes were generated from cDNA instead of gDNA. cDNA probes were generated from RNA from asynchronous cultures containing mixed stages for IT and 3 culture-adapted Kilifi patient isolates, 10668, 9106 and 8383. 5ug of the same RNA sample used to generate cDNA for the probes was resolved on a 1.5% agarose gel and transferred to a nylon membrane. The specific probes were hybridised sequentially starting with the IT cDNA probe, then followed with the wild isolates probes10668, 9106 and 8383. Using the specific probes, a band slighlty under 2kb, corresponding to *rif* transcript was detected in homologous RNA (figure 4.16). For IT an intense band was observed just below the 18s rRNA band (Figure 4.16. A). For 10668 a faint band was detected just below the 18s rRNA band (Figure 4.16. B, lane 2). The faint band was masked by the rRNA band and was only clearly visible after a long exposure (Figure 4.16. B (II) lane 2). No *rif* transcript signal was detected in 9106 and 8383 using homologous cDNA probes (Figure 4.16. C and D), but this was probably due to masking by the rRNA signal. Non-specific bands were observed below the 28s rRNA band at about 3kb and 2.7kb and below the *rif* band at about 1.5kb, especially after probing with 10668 cDNA probe.

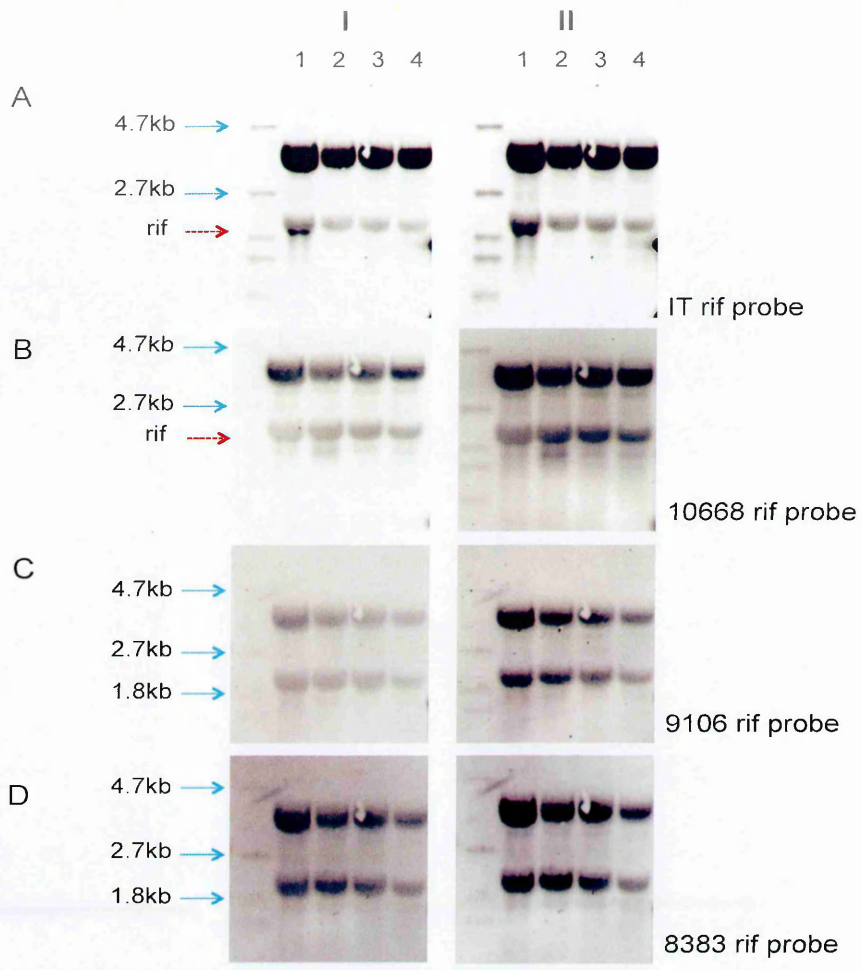


Figure 4.16: *Northern blots of 5ug total RNA from pooled timepoints from IT (1), 10668 (2), 8383 (3) and 9106 (4) hybridised with probes generated from IT (A), 10668 (B), 9106 (C) and 8383 (D) rif RT-PCR products sequentially, after 20 minutes exposure (A) and 60 minutes exposure (B). Red arrow points to the ~2kb rif band in IT RNA (lane 1) of the blot probed with rif IT cDNA probe (A) and in 10668 RNA (lane 2) probed with 10668 rif cDNA probe (B). No clear signal was detected for 9106 and 8383 using homologous cDNA probes. L: DIG-labbeled RNA marker I (Roche).*

4.6. Discussion.

The high sensitivity of RT-PCR in detecting transcripts as compared to northern blotting enabled detection of multiple *rif* and *stevor* transcripts throughout the IDC of laboratory and wild isolates. Despite inherent limitations of our approach, which include primer and cloning bias, we were able to observe high intra- and inter-isolate *rif* transcript diversity. Previous studies have reported transcription of multiple *rif* and *stevor* at different IDC stages (Otto et al., 2010, Daily et al., 2005, Rovira-Graells et al., 2012).

The lack of clear dominant *rif* transcripts at certain time points especially in 3D7 and the field isolates could point to limited sampling of transcripts. In order to clearly define transcription profiles much more sampling depth using high throughput sequencing technologies will be required. Despite this shortcoming our study was able to reveal patterns of *rif* and stevor transcription through the asexual blood stage cycle. Clear stage-specific patterns of *rif* expression emerged in laboratory and culture-adapted wild isolates. In some wild isolates dominant *rif* were observed at certain stages, especially in the schizont stage.

It was interesting to recover an UpsA var-linked *rif* PF13_0004 in our 3D7 line at ring stage (RIF_K0044). Previous published expression data in 3D7 parasites showed that in 3D7 parasites selected on immune sera or for HBEC-binding group A *var* were preferentially expressed and so were the *rif* neighbouring them (Wang et al., 2009, Claessens et al., 2012).

In a previous study, Bachmann et al observed the same genes in the two peaks of *rif* expression, at ring and trophozoite (Bachmann et al., 2012) but this could be due primer bias and/or sequence homology in the target region they analysed. In our study most isolates transcribed different *rif*s at the different IDC stages. However we did observe similar *rifs* coming up in rings and schizonts. This was observed in IT, SA075 R+/ R-, 10594, 10761 and 10814, although in 10814 the shared transcript was observed in late ring/ early trophozoites and schizonts. In addition we observed less transcript diversity with a clear dominant *rif* transcripts in schizonts in most isolates. Although we did observe dominant schizont *rif* in 10594 and 10595, interpretation of expression profiles in the 2 isolates was complicated by suspected gDNA contamination of cDNA. This schizont-associated transcript could be the merozoites transcript, and hence carried over to the rings after invasion. A previous study reported two *rif* that were expressed at the beginning and end of the 48 hour IDC in 3D7 (Wang et al., 2009) suggesting that these *rifs* played a role in the early and late stages of the IDC. We detected one of the two *rifs* PF13_0006 in late trophozoites in our 3D7. We did not detect it in schizonts probably due to under-sampling.

Interesting *rif/stevor* expression profiles were observed in the rosetting culture-adapted wild isolate SA075. There was a clear difference in *rif* and stevor transcripts at trophozoite stages in the rosetting clone and the non-rosetting fraction. A previous study showed two *rifs* upregulated in rosette-selected IT (Claessens et al., 2011). One of the upregulated *rifs* termed IT4rifA_054 was head-to-head with the most upregulated *var*, the rosette associated ITvar9. This *rif* was mostly expressed in the mature pigmented

trophozoites (28hpi). The other upregulated *rif* IT4rifA_042 was predominantly expressed in rings (12hpi). Our IT line that underwent rosette-selection had a very restricted profile, similar to the Kisumu rosette-selected lines. This might suggests that selection pressure does play a role in *rif* and stevor expression. Further analysis is needed to link specific *rif* and *stevor* with rosetting phenotype.

Stevor expression profiling was significantly limited by primer bias. Our primers were only able to pick up 59 *stevor* variants out of a possible 160 variants from all four wild isolates. The observed high level of sequence sharing in Kilifi isolates could be a true phenomenon or could be due to primer bias. For a clearer picture of *stevor* transcription different primers sets with better repertoire coverage will be required. Despite the substantial bias, we observed very little sharing of *stevor* transcripts in Kilifi and Kisumu isolates. In fact only two *stevor* were shared between Kilifi and Kisumu isolates. A study looking at repertoire diversity in Kilifi and laboratory isolates detected shared stevor in wild and laboratory parasites (Blythe et al., 2009).

One major limitation was the inability to capture all *rif* and *stevor* variants using the primer sets. Due to the high variability of these multigene families, it is impossible to design universal primers that will capture the entire repertoire in wild isolates. The primers used were designed on laboratory and

In conclusion, *rif* and *stevor* expression profiling was evaluated in laboratory and wild isolates across the IDC. Clear stage specific and isolate specific expression patterns were revealed. Dominant transcripts were observed in specific stages in certain parasites.

Conserved *rif* and *stevor* genes were observed in Kilifi isolates, with little sharing between laboratory and wild isolates. Most interesting was identification of rosette-associated *rif* and *stevor* transcripts in a patient isolate.

In the northern blotting analysis all probes labelled the 2.1kb and 4.1kb non-specific rRNA even in the low concentration (1ug) RNA sample, suggesting that the signal was not probe specific therefore migh be from the anti-DIG antibody, from non-specific products amplified during PCR for probe preperation, from the gene-specific primers or the T7 promoter sequence. This signal was very high, and changing the stringency of washes and hybridisation at higher temperatures did not reduce the intensity. These highly abundant RNA species gave intense bands at 2kb (18s) and 4kb (28s) that interfered with correctly spliced *rif* transcript signal which is of about the same size as the 18s rRNA, and a previously detected unspliced product of about the same size as the 28s rRNA (4kb, [Kyes, unpublished]). BLAST analysis of the complex *rif* probe against NCBI nucleotide database identified the T7 polymerase anchor sequence included in all the primer sets as a region likely to hybridise to rRNA. In addition, the complex primers used to amplify *rif* for probe generation did non-specifically amplify non-target genes in *P. falciparum* including rRNA and ring-infected erythrocyte surface antigen (RESA). The non-specific band at about 3kb is most likely RESA (transcript length 3.2kb).

Overall full-length, sense *rif* transcripts were detected in IT and in a culture-adapted Kilifi isolate, 10668. This implies that of the many transcripts detected by RT-PCR some are correctly spilced and in the correct orientation. This is in line with literature showing

multiple *rif* variants transcribed in an isolate or cloned parasite. Post transcription regulation may play an importatnt role in regulating *rif* expression. The observation that only homologous probe from cDNA was able to detect *rif* transcript suggests that *rif* expression is isolate specific, at least in IT and the three Kilifi isolates, as no signal was detected using probes from heterologous isolates. This would also suggest that a small fraction of the genomic repertoire is expressed since the gDNA probes were not able to detect transcripts.

An important and missing piece of information is whether all the detected transcripts are made into protein during the parasite's IDC. Reagents to analyse protein localisation for *rif* and *stevor* in wild isolates were not available at the time of our study and that was a major limitation. Earlier studies have detected RIFIN and STEVOR proteins in multiple stages in laboratory isolates. It is unclear whether all the upregulated transcripts detected were properly spliced and made into protein. This could be addressed in future work using recombinant proteins and RIFIN/STEVOR-specific antisera.

CHAPTER 5

5. Analysing *rif* and *stevor* expression profiles by 454 amplicon sequencing.

5.1. Introduction.

During its life cycle *P. falciparum* undergoes different selection pressures as it matures through different morphological stages and in different hosts and host environments. Switching between different variant types of VSA was linked to switches in adhesive phenotypes in earlier studies (Roberts et al., 1992). This function has been well described for *var* genes, where a switch in *var* gene expression correlated with a change in binding phenotype of the parasite to ICAM1 (Smith et al., 1995) and HBEC 5i cell line (Claessens et al., 2012). Expression switching under different selection pressures has not been well documented for *pir* genes. Members of the *pir* multigene families have been observed in all *Plasmodium* species, unlike *var* genes that are only found in *P. falciparum* and *P. rechenowi.* It is therefore likely that the antigens encoded by *pir* genes may play a role in cytoadhesion and in disease pathogenesis in other *Plasmodium* species. Recently a link between parasite virulence and expression of subsets of the *pir* gene family has been shown in *P. chabaudi* (Spence et al., 2013). In the study it was observed that mosquito transmitted

(MT) *P. c. chabaudi* AS parasites were less virulent than serially blood passaged (SBP) parasites, because of a modified immune response to the blood stage parasite. The main difference between the MT and SBP parasites was in *cir* (*pir* genes in *P. chabaudi*) gene expression, with the SBP parasites expressing a dominant *cir*, while the MT had increased but diverse *cir* expression (Spence et al., 2013). This study revealed mosquito transmission directly modified blood stage parasites, by modifying expression of

virulence determinants, *cir* genes, and thus modifying the host immune response, which in turn regulates virulence.

Although many studies have looked at *var* expression and regulation there is poor understanding of the expression patterns and mechanisms that regulate *rif* and *stevor* gene expression. *rif* and *stevor* are expressed in multiple stages of the parasite life cycle and in many stages of the IDC. In addition, these genes undergo switching thus the antigens are clonally variant (Fernandez et al., 1999, Lavazec et al., 2007, Niang et al., 2009). These data suggest that *rif* and *stevor* are under strict transcriptional control and are upregulated in stages in which they need to functional. Studies on temporal expression of *rif* and *stevor* have been limited to laboratory isolates (Otto et al., 2010, Bozdech et al., 2003, Bachmann et al., 2012, Claessens et al., 2011, Llinas et al., 2006) with limited data on expression in field isolates. One recent study looking at VSA transcription in a limited number of wild isolates by realtime PCR and amplicon cloning and sequencing (Bachmann et al., 2012) showed two peaks of *rif* and *stevor* expression, in the early and late IDC stages, with multiple *rif* and *stevor* variants being transcribed. Other studies looking at *rif* and *stevor* transcription revealed upregulation in wild isolates as compared to laboratory adapted lines (Lemieux et al., 2009, Blythe et al., 2008).

Even more limited is data on the effect of selection pressure on *rif* and *stevor* expression. In a study looking at the effect of rosette-selection on VSA expression in IT one *var* ITvar9 was highly upregulated in the rosetting line (Claessens et al., 2011). Although many *rif* and *stevor* were expressed above background only one specific UpsA *rif*

neighbouring the dominant *var*, ITvar9 was significantly upregulated in the rosetting line. This suggested transcription co-regulation of UpsA *var* and their neighbouring *rifA*. *Stevor* were not differentially expressed in the rosette-selected parasites. Another study by the same group looked at effect of selection for binding to HBEC-5i cells on VSA expression (Claessens et al., 2012). Group A *var* genes were upregulated in the selected lines in 3D7, IT and HB3. The UpsA *var*-linked *rif* neighbouring the upregulated *var* were also upregulated in 3D7 and HB3 suggesting an association in transcription between the two VSA gene families. The link in expression between *var* and *rif* has previously been reported in a study showing that parasites selected to express PfEMP1 associated with severe malaria upregulated UpsA var as well as the neighbouring rifA1 (Wang et al., 2009).

Elucidation of the role of RIFIN and STEVOR will require better characterisation of expression patterns through the different parasite stages and under different selection pressures. Our previous attempt to analyse *rif* and *stevor* expression profiles by cloning and capillary sequencing described in Chapter 4 was hampered by limitation in repertoire coverage due to the few numbers of clones analysed as well as cloning bias. Significant primer bias was noted for *stevor*. The results therefore did not represent an accurate depiction of expression patterns across the parasite IDC. Clone-free 454 amplicon sequencing was carried out to overcome cloning bias, increase *rif* and *stevor* repertoire coverage and to validate the results previously obtained from cloning and sequencing *rif* and *stevor* amplicons. Additional wild isolates put under different selection pressures

were analysed to understand whether selection had an effect on *rif* and *stevor* transcription.

454 sequencing from Roche 454 Life Sciences™ is a pyrosequencing system where a single strand of a DNA fragment is immobilized on a micron-sized bead, amplified in an emulsion and sequenced via sequencing-by-synthesis chemistry. 454 sequencing technology allows for massively parallel sequencing of amplicons therefore help to resolve diversity of the complex patterns of *pir* expression seen with capillary Sanger sequencing, just by increasing the depth of sequencing. Massively parallel sequencing of the targeted regions (rif/stevor) would permit better quantification of reads of different variant types and thus enable quantification of transcripts and identify dominant transcripts, more reliably than Sanger sequencing. Important advances in 454 technology allowed longer read lengths than other next generation sequencing platforms, and therefore might permit sequencing of the 850-900 bp *rif* amplicons and ~550-650 bp *stevor* amplicons (Figure 5.2), which are necessarily long in order to encompass semi-conserved regions where the degenerate primers can be designed and the hypervariable central regions to characterize variant-specific expression profiles. The GS FLX Titanium XL+ that can achieve read length of up to 1kb was used. With the ability to multiplex samples, over two hundred *rif* and *stevor* amplicons were sequenced in parallel.

5.2. Objectives

Deep sampling of *rif* and *stevor* genomic repertoires and expression profiles by 454 amplicon sequencing to:

i.) Optimise 454 sequencing and data analysis protocol for *rif* and *stevor* amplicon sequencing.

ii.) Analyse *rif* and *stevor* expression across IDC stages in laboratory isolates.

iii.) Analyse *rif* and *stevor* expression across the IDC in patient isolates under different selection pressures.

5.3. Method.

The methods in this chapter have been detailed in Chapter 2 (materials and methods) and Chapter 4 (methods). *Rif* and *stevor* expression analysis was carried out in laboratory and wild isolates. Expression profiles were analysed at 6 time points in laboratory isolates, and 4 to 7 time points in wild isolates. There was a total of 16 isolates, 3 laboratory isolates (3D7, IT and HB3), 12 Kilifi isolates, 3 from patients with severe malaria and 9 from patients with non-severe malaria, and 1 Kisumu isolate from a patient with severe malarial anemia. Effects of selection pressure on *rif* and *stevor* expression was analysed by comparing *rif/stevor* transcripts obtained from different lines of wild isolates that had been exposed to different selection pressures; host immune pressure, culture-adaptation, and binding phenotypes (Table 5.1 and Figure 5.1). To analyse the effect of host immune pressure *rif* and *stevor* expression profiling was done on fresh-from-the-arm and ex vivo parasites. Culture-adapted wild isolates, as well as isolates selected for different binding phenotypes were obtained from an on-going *var* study (Dr. Pete Bull). Two Kilifi isolates, 9106 and 10668 had been selected for binding to human brain endothelial cell line and one Kisumu isolate, SA075 had been selected for rosetting. The rosetting fraction of SA075 was cloned by limiting dilution. The culture-adapted unselected and selected lines were analysed to determine whether there was an effect of selection on *rif* and *stevor* expression (Appendix Table 9.6).

As previously described in Chapter 4 RNA was extracted from the different parasite lines at different time points. For the laboratory isolates 3D7, IT and HB3, RNA was extracted at 6 time points, corresponding to early ring (ER), late ring (LR), early pigmented

trophozoite (ET), mid trophozoite (MT), late trophozoite (LT) and schizont (S) stages. For patient isolates, RNA was extracted from fresh-from-the-arm ring samples (R) and from ex vivo 24 hour cultured trophozoite samples (T). For the culture-adapted Kisumu and Kilifi lines RNA was extracted at 4 time points, ring (R), late ring to early pigmented trophozoite (LRET), mid to late trophozoite (T) and schizont (S) (Appendix table 9.5). For the Kilifi ex vivo isolates the time course experiment was carried out by culturing the ex vivo isolates for a single IDC. RNA was extracted at 7 time points every 10 hours across the IDC (Appendix table 9.5). Genomic DNA (gDNA) was extracted from asynchronous cultures at trophozoite stage.

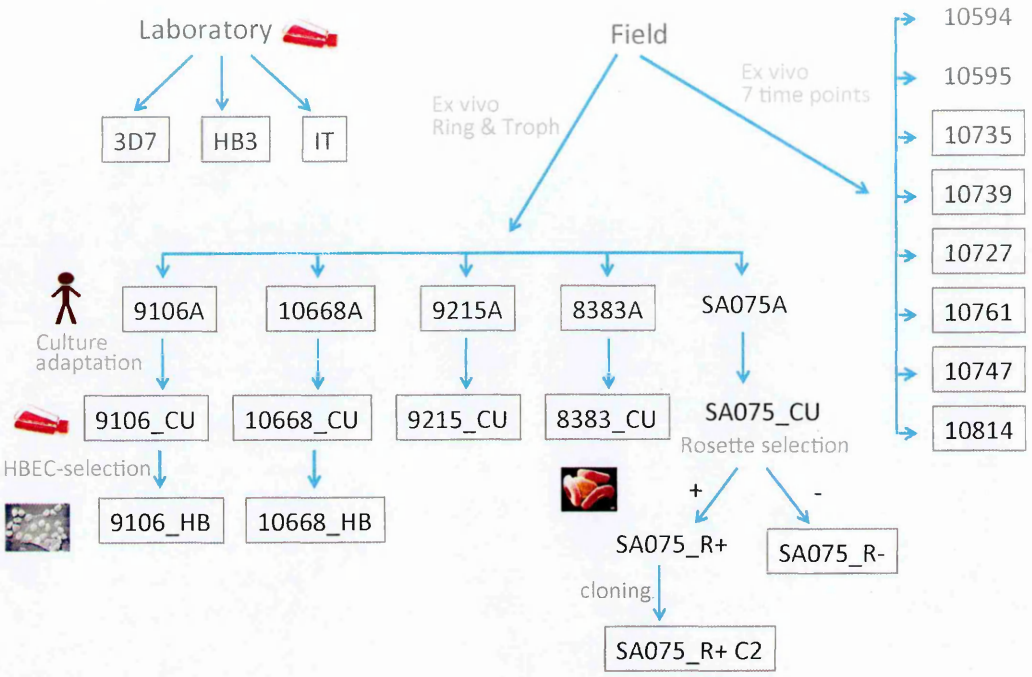


Figure 5.1. *Isolates used for rif and stevor expression profiling. Laboratory lines: 3D7, HB3 and IT. Wild isolates: 9106, 10668, 9215, 8383, SA075, 10594, 10595, 10735, 10739, 10727, 10761, 10747 and 10814. A subset of the wild isolates were culture adapted: 9106, 10668, 9215, 8383 and SA075. A subset of the wild-adapted parasites were selected for binding to HBEC (9106_HB and 10668_HB) and rosetting (SA075_R+ and SA075_R-). The rosetting fraction of SA075 was cloned by limiting dilution to give SA075_R+_C2. Parasite lines that were used in this experiment are enclosed in boxes.*

PCR amplification of *rif* and *stevor* was carried out using gene family-specific degenerate primers (Figure 5.2 and Table 2.1). For *rif* previously published complex primer sets with 3 forward and one reverse primer were used (Kyes et al., 1999). The reverse primer was modified to correct an error in the published primer sequence and to make it more degenerate. For *stevor* a newly designed forward primer was used together with a previously published reverse primer (Albrecht et al., 2006). The amplicons were between 850 and 950 base pairs (bp) for *rif* and 550 to 650bp for *stevor* (Figure 5.2).

Amplicons were purified and shipped to the Sanger institute for 454 sequencing. Library preparation was carried out according to the GS FLX Titanium Rapid Library Preparation Kit (Roche) protocol and sequencing done on the GS FLX+ (Roche) (chapter 2 section 2.8). Prior to sequencing amplicons were multiplexed and pooled into 6 pools of 48 amplicons per pool. A total of 246 amplicons were sequenced, 136 *rif* and 110 *stevor* amplicons.

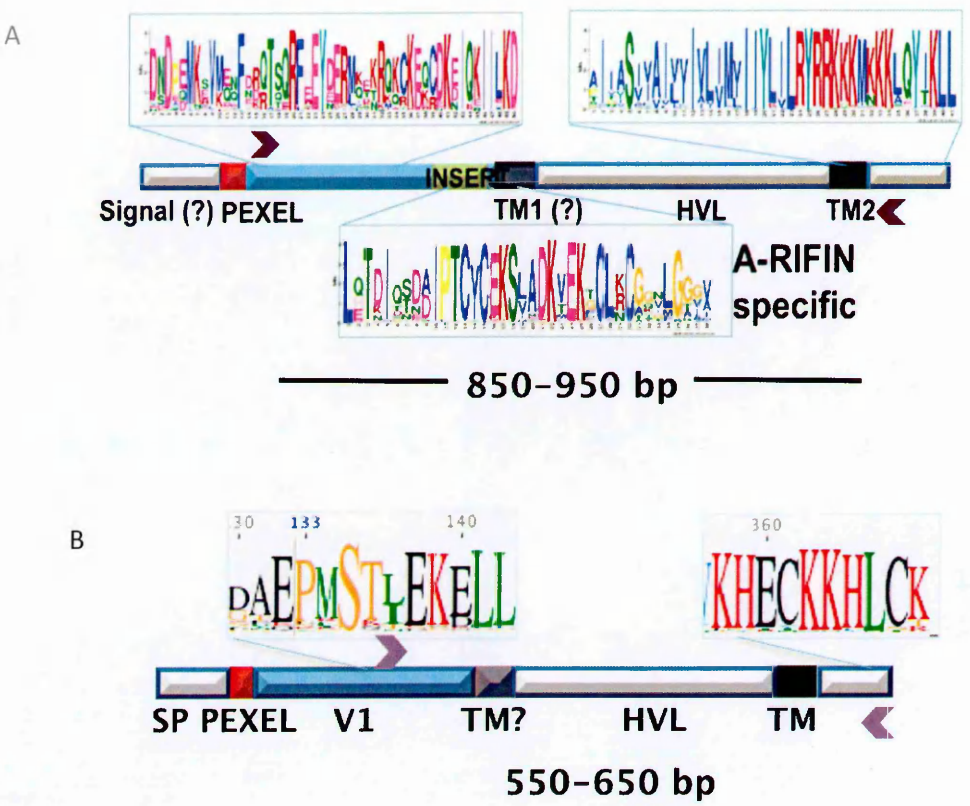


Figure 5.2. Schematic diagram of rif and stevor genes showing the gene features and primer positions (purple arrow heads), and the length of amplicons, 850-950bp for rif and 550-650 for stevor. TM: transmembrane domain, HVL: hypervariable loop, SP: signal peptide, V1: variable region 1.

5.3.1. Sequence analysis.

For the 3 laboratory lines (3D7, IT and HB3) and 4 wild isolates (9106, 9215, 10668 and SA075) whole genome sequence was available (Table 5.1). For one Kilifi isolate the genome sequence data was very poor and could not be used for further analysis (Table 5.1). Genome annotation for the wild isolates was done automatically with no manual curation. Previously identified motifs unique to *pir* gene family (discussed in chapter 3) were used to annotate *pir* genes in the Kilifi parasites genomes (work done By Dr. Thomas Otto). This annotation was automated therefore a few *pir* genes were missing annotations or incorrectly annotated, and some had incorrect gene models requiring a lot of manual curation before further analysis.

Reads generated by 454 sequencing were mapped against the homologous genome where available. Raw reads were extracted from the 454 SFF files and a Perl script was used to group the reads by their barcode (multiplex identifiers). Reads of each bar code were mapped against their reference genome, using SMALT (ftp://ftp.sanger.ac.uk/pub/resources/software/smalt/), with the following parameters; k-mer length of 13 and a step size of 3 (index -k 13 -s 3), random placing of reads (mapping -r 0), 50% minimum identity of complete read and region in the genome (-y 0.5). The resulting output (sam file) was transformed with samtools (Li et al., 2009) to a bam file. The bam file was loaded onto Artemis (release 15.0.11) (Rutherford et al., 2000). In Artemis BAM view (Carver et al., 2013) RPKM (reads per kilo base pair per million) and read counts were calculated for all the annotated *rif* and *stevor* genes. For transcript abundance read counts were used because the amplicon lengths were similar hence there was no need to normalise for gene lengths by using RPKM.

Table 5.1 *List of Isolates used for rif and stevor expression profiling by cloning and capillary sequencing and 454 sequencing of total amplicon.*

Parasite isolate	Clinical phenotype of original isolate.	Parasite lines studied for *pir* expression during IDC	IDC stages of RNA sampling	Whole genome sequence	Amplicon sequencing method	
					Capillary sequencing	454 sequencing
3D7	N/A	3D7	6 time points, ER, LR, ET, MT, LT, S of the IDC of the laboratory isolates	Y	Y [r]	Y [r, s]
IT	N/A	IT		Y	Y [r]	Y [r, s]
HB3	N/A	HB3		Y	-	Y [r, s]
8383	NS	8383A	*ex vivo* ring & trophozoite	-	-	Y [r, s]
		8383_CU	4 time points, R, P, T and S of the Culture-adapted line	Y[a]	-	Y [r, s]
9106	IC&RD	9106A	*ex vivo* ring & trophozoite	-	-	Y [r, s]
		9106_CU	4 time points, R, P, T and S of the Culture-adapted line	Y	-	Y [r, s]
		9106_HB	4 time points, R, P, T and S of the Culture-adapted and HBEC-selected line	-	-	Y [r, s]
9215	RD	9215A	*ex vivo* ring & trophozoite	-	-	Y [r, s]
		9215_CU	4 time points, R, P, T and S of the Culture-adapted line	Y	-	Y [r, s]
10668	IC&RD	10668A	*ex vivo* ring & trophozoite	-	-	Y [r, s]
		10668_CU	4 time points, R, P, T and S of the Culture-adapted line	Y	-	Y [r, s]
		10668_HB	4 time points, R, P, T and S of the Culture-adapted and HBEC-selected line	-	-	Y [r, s]
SA075	SMA	SA075_R+ (clone)	4 time points, R, P, T and S of the Culture-adapted and rosetting clone	-	Y [r, s]	Y [r, s]
		SA075_R- (uncloned fraction)	4 time points, R, P, T and S of the Culture-adapted rosetting negative uncloned fraction.	Y	Y [r, s]	Y [r, s]
10735	NS	10735	7 time points of *ex vivo* wild isolates (per isolate parasite stage in Appendix table 1)	-	Y [r]	Y [r]
10739	NS	10739		-	Y [r, s]	Y [r]
10727	NS	10727		-	-	Y [r, s]
10747	NS	10747		-	-	Y [r, s]
10761	NS	10761		-	Y [r, s]	Y [r, s]
10814	NS	10814		-	Y [r, s]	Y [r, s]

The wild isolates were collected from children with either non-severe malaria (NS) or severe malaria with respiratory distress (RD), respiratory distress with impaired consciousness (IC&RD) or severe malaria anaemia (SMA). For the time series expression profiling RNA was sampled at 6 IDC stages in laboratory isolates; early ring (ER), late ring (LR), early trophozoite (ET), mid trophozoite (MT), late trophozoite (LT) and schizont (S), 2 stages, ex vivo ring (R) and trophozoite (T), 4 stages in cultured and selected lines; ring (R), late ring to early pigmented trophozoite (P), mid and late trophozoite (T) and schizont (S). For laboratory isolates and 5 of the culture-adapted wild isolates whole genome sequencing data was available (Y). Y^{a}: poor quality whole genome sequence data for Kilifi isolate 8383. For some isolates pir expression profiling was done using both capillary and 454 amplicon sequencing; Y: sequence data available, -: no sequencing data. r: rif, s: stevor. Complete description of isolates used in this study in Appendix table 9.6.

5.4. Results

5.4.1. Optimising 454 sequencing runs.

The first two 454 sequencing runs (4 lanes each with 48 multiplexed amplicons) resulted in very short mean read lengths, about 100 base pairs (bp), much shorter than the expected amplicon length of 850-950bp for *rif* and 550-650bp for *stevor* (Figure 5.2). Sequence analysis revealed that the short reads were primer-primer concatemers. This was as a result of amplicon contamination with primer dimers. The amplicons had passed quality control checks and the contaminating primer dimers were not visible on Agarose gels (Figure5.4. I) or on the Agilent Bioanalyser traces (Figure 5.4. II), but they were in sufficient quantities to significantly interfere with sequencing resulting in suboptimal read lengths (Figure 5.3). The short fragments (primer dimers) were preferentially amplified in the emulsion PCR (Figure 5.5) and therefore preferentially sequenced. This resulted to recovery of short reads from the sequencing runs.

To overcome this problem the pools of amplicon were taken through high stringency solid phase reversible immobilization (SPRI) beads clean up, using a ratio of 0.7 SPRI beads to DNA solution instead of the normal ratio of 1.8 and resequencing of the libraries in a third 454 sequencing run. Much longer reads were recovered in the third sequencing attempt with the *rif* pools still containing a significant number of short reads (Figure 5.3).

Because of limited resources and samples it was not possible to repeat sequencing for all the pools. Therefore only *rif* pools 1 and 2 and stevor pools 5 and 6 were repeated. *Stevor* pool 4 had succeeded in the first run (Figure 5.3). Low quality data was obtained from the "failed" sequenced *rif* pool 3. Because the aggressive cleanup was

done on the pools and not individual samples there was sample-to-sample variation in reads even within the same pool.

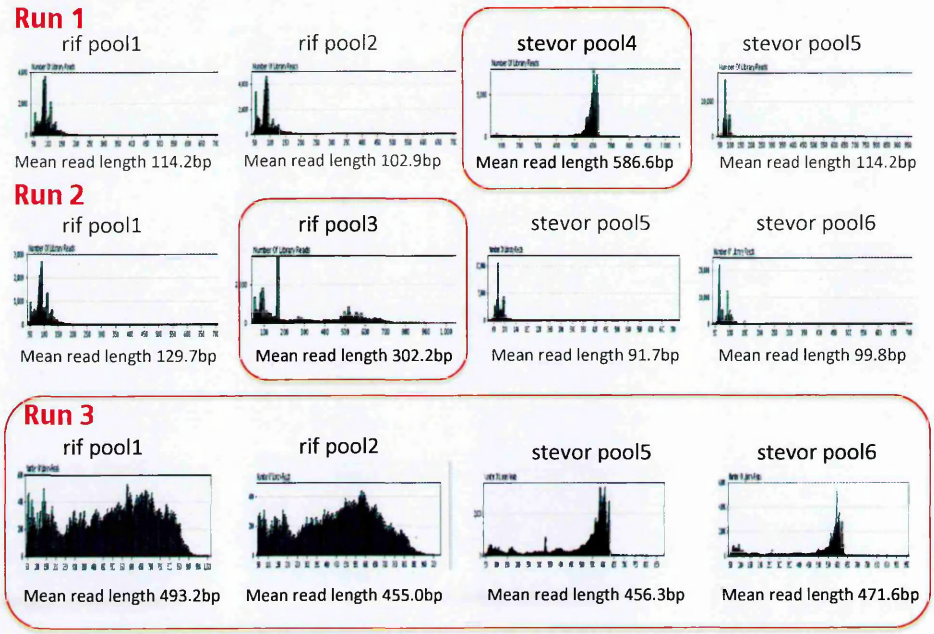


Figure 5.3. *Read lengths distributed in the 2 failed 454 runs (1 and 2) and the 3rd and successful run. Graphs show the number of reads (Y axis) plotted against the length in base pairs (X axis) for all the four pools per run in 3 sequencing runs. The mean read length in base pairs for each pool is indicated below each graph. Red boxes are drawn around pools that had longer reads (mean read length >100bp) that were used for subsequent analysis.*

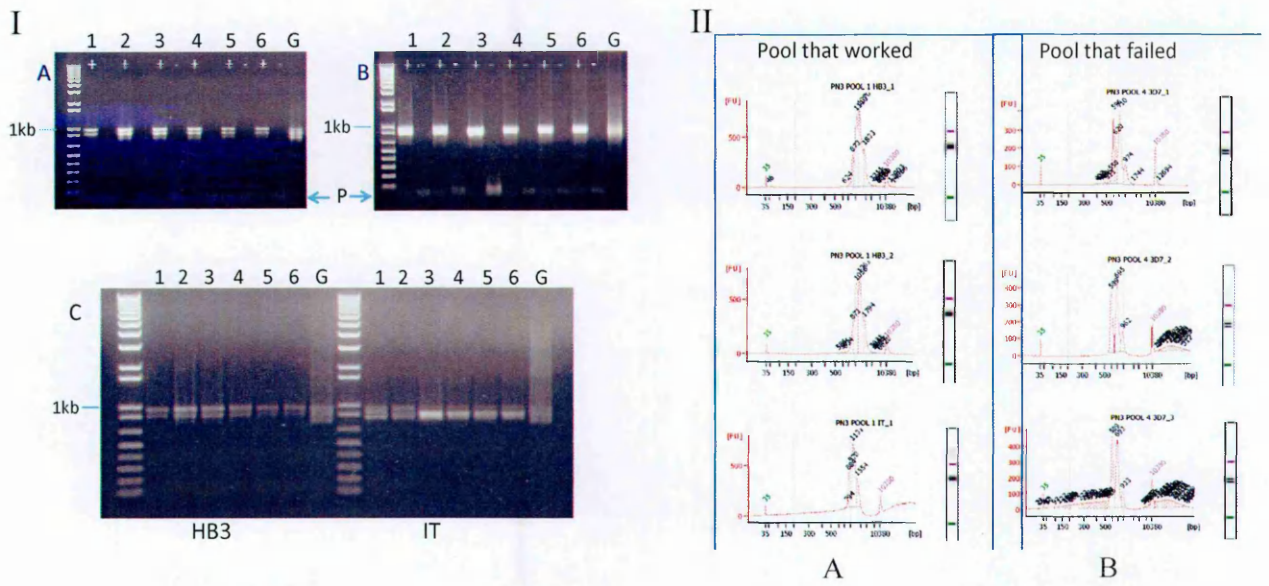


Figure 5.4. *rif amplicons detected on 1% agarose gels and analysed on Agilent Bioanalyser. (1) Agarose gel images showing rif amplicons from laboratory isolates HB3 and IT (A and B) at 6 different time points and gDNA prior to PCR purification and after purification (C). (+) and (-) indicate samples with (RT+) and without (RT-) the reverse transcriptase enzyme. (2) Agilent Bioanalyser traces of libraries prior to pooling and sequencing. A) Libraries in a pool that yielded long reads and B) libraries in a pool that yielded short sequences. The Y-axis indicates the DNA quantity (fluorescence units) and X-axis indicates the length in bp. The 2 peaks at the begging (35bp) and the end (10380bp) indicate lower and*

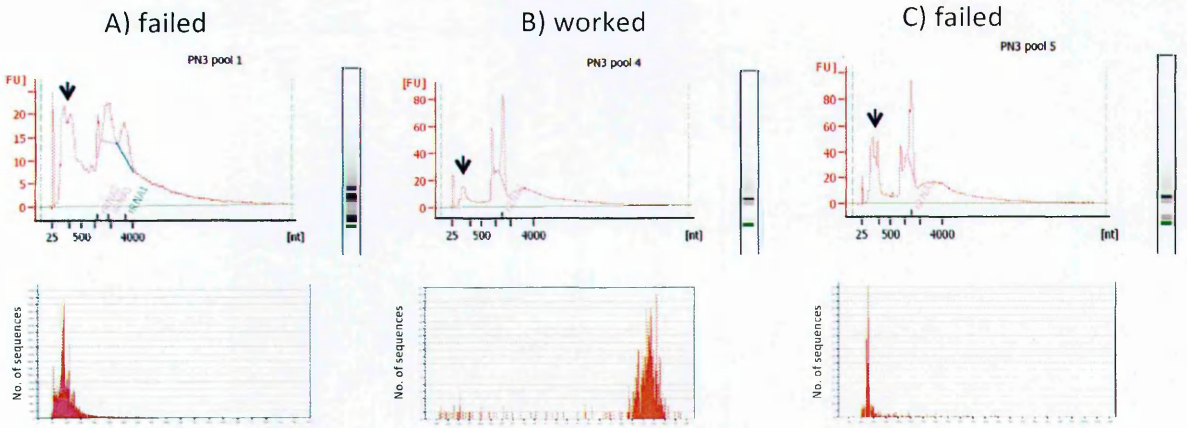


***Figure 5.5**. High abundance of short fragments in pools A and C after emulsion PCR resulted in recovery of short sequence reads. Upper panel Agilent Bioanalyser traces of pools after emulsion PCR. Y-axis is the DNA quantity (fluorescence units) and the X-axis is the length in base pairs (bp). Arrows point to the peak made up of short DNA fragments as a result of contamination of amplicons with primer dimers. The larger peaks above 500bp are the target amplicons. Lower panel shows the distribution of read lengths from 454 sequencing of the pools in the upper panel. Y-axis is the number of reads and the X-axis is the length in base pairs (bp). Pool A and C had larger proportions of small fragments and resulted in failed sequencing runs characterised by short reads. Pool B had smaller proportion of contaminating fragments in the amplicons and this pool yielded longer sequence reads.*

5.4.2. Optimising 454 sequence data analysing protocol.

Due to the suboptimal 454 sequencing results, many reads did not span the entire amplicon, more so for the longer *rif* amplicon (Figure 5.1). This fact compounded by the large number of reads (raw reads mapping data on tables 9.8-9.12) required approaches to analyse the data. Two analysis approaches were taken, analysis by mapping reads on to corresponding genomes and de novo assembly of reads into contigs. The mapping approach was used for parasite isolates with whole genome sequence data available (Table 5.1) that is 3 laboratory isolates 3D7, IT and HB3, and four wild isolates 9106, 9215, 10668 and SA075. Reads were mapped onto corresponding genomes as described in the methods above and the number of reads that map to features annotated as *rif* or *stevor* was visualised and computed in Artemis. De novo assembly of reads into contigs was used to assemble reads where mapping to a reference genome was not possible due to unavailable or poor whole genome sequence data (Table 5.1). The results of the de novo assembly approach are discussed in chapter 6.

5.4.3. Confirmation and correction of genome annotation by mapping reads to genomes.

Whole genome sequence data from laboratory isolates has undergone extensively manual curation, and it was possible to extract complete *pir* repertoires (from 3D7) and correct annotations for the annotated partial *pir* repertoire in IT and HB3. The wild isolates on the other hand had undergone automated curation and for some the *pir* repertoire was incomplete (Table 5.2). For some genomes failed sequencing in

long tracts of the genome resulted in very few *pirs* annotated, with some genomes having as few as 17 annotated *rifs* out of an expected 180, and 2 annotated *stevors* out of an expected 40 (Table 5.2). For some genomes misannotations were observed where non-*pir* sequences were annotated as *pir*, and a few incorrectly spliced *pir* genes. Manual correction was done for isolates used in the 454 amplicon sequencing study. The *pir* genes with missing annotations were identified by mapping all 454 sequence reads to the genome, and scanning by eye for reads mapping on regions with no annotations. This way most missing annotations were corrected for *pir* genes amplified by the degenerate primer sets. Some *pirs* mostly those not recovered in the 454 sequencing data. For one wild isolate genome, SA075 where many *rif* and *stevor* were not annotated, by mapping all reads from SA075 sequenced amplicons an additional 55 *rifs* and 9 *stevors* were annotated For IT version 2 of the genome, the available version for mapping at the time of this analysis, had many annotation errors. The revised version IT version 3 had fewer *pirs* annotated, 132 *rifs* and 39 *stevors* (as compared to 138 *rifs* and 43 *stevors in* IT v2). The version 3 was used to correct version 2 for misannotations and improper splicing.

Table 5.2. *Number of pirs from whole genome sequences from laboratory and field isolates.*

Parasite isolate	Whole genome sequence	Complete or partial *pir* counts		*pir* expression data
		No. of *rif*	No. of *stevor*	
3D7	Y	184 (complete)	42 (complete)	Y
IT	Y	138	43	Y
HB3	Y	142	35	Y
6816	Y	177	45	N
8383	Y[a]	17	3	Y
9106	Y	155	42	Y
9215	Y	168	39	Y
9605	Y	177	41	N
9626	Y[a]	19	2	N
9775	Y[a]	35	9	N
10668	Y	179	40	Y
10936	Y	181	41	N
10975	Y[a]	27	9	N
11014	Y[a]	17	2	N
11019	Y	197	48	N
SA075	Y[a]	95#	31#	Y

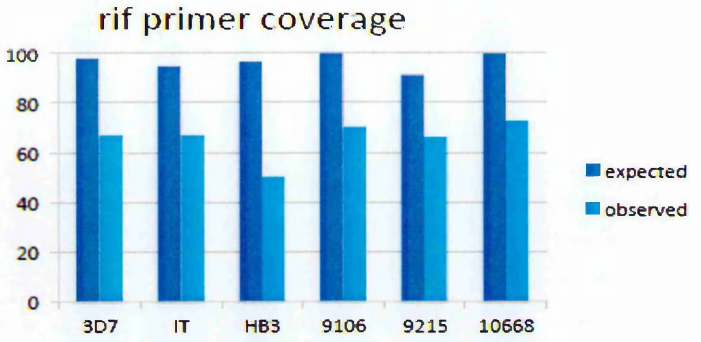
a - noticeably incomplete pir repertoire

 # - additional 55 rifs and 9 stevors annotated

5.4.4. Primer coverage

Rif and *stevor* primer coverage was estimated using reads from gDNA for all the isolates that have whole genome sequence data (Table 9.6). In view of the incomplete annotation for some of the Kilifi genomes (see Table 5.2) an assumption was made that the repertoires were similar to the 3D7 reference with 184 *rif* and 42 *stevor* copies per genome. Using these estimates a normal distribution was computed based on the total number of reads from gDNA that mapped to *rif* or *stevor* on corresponding genomes. *rif* primer coverage varied from 50% in HB3 to 72% in 10668 (Figure 5.6. A). In 3D7 *rif* primer coverage was 67% based on reads from gDNA and 81% based on reads from both gDNA and cDNA. *stevor* primer coverage varied from 47% in HB3 to 72% in 3D7 and IT (Figure 5.6. B). Of importance was the inability of stevor primers to capture any members of the more divergent minor group (discussed in Chapter 3) (Blythe et al., 2009, Schreiber et al., 2008). Overall the observed primer coverage was significantly less than the expected coverage for all the isolates, an inherent problem in trying to capture all members of highly variable multigene families using degenerate primers. Thus the resulting expression profiles presented are for a subset of *pir* genes as it was not possible to design primers to capture the complete repertoire. An ideal strategy that was not feasible at the time of this study would be whole transcriptome sequencing.

A



B

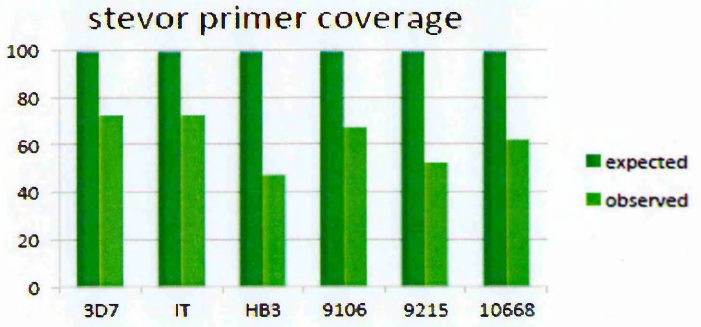


Figure 5.6. *rif (A) and stevor (B) primer coverage based on reads from gDNA obtained by 454 amplicon sequencing. Y-axis shows rif/ stevor percent expected or observed coverage computed from number of reads recovered from sequencing gDNA amplicons for the different isolates. X-axis represents isolates analysed by mapping reads to homologous genomes.*

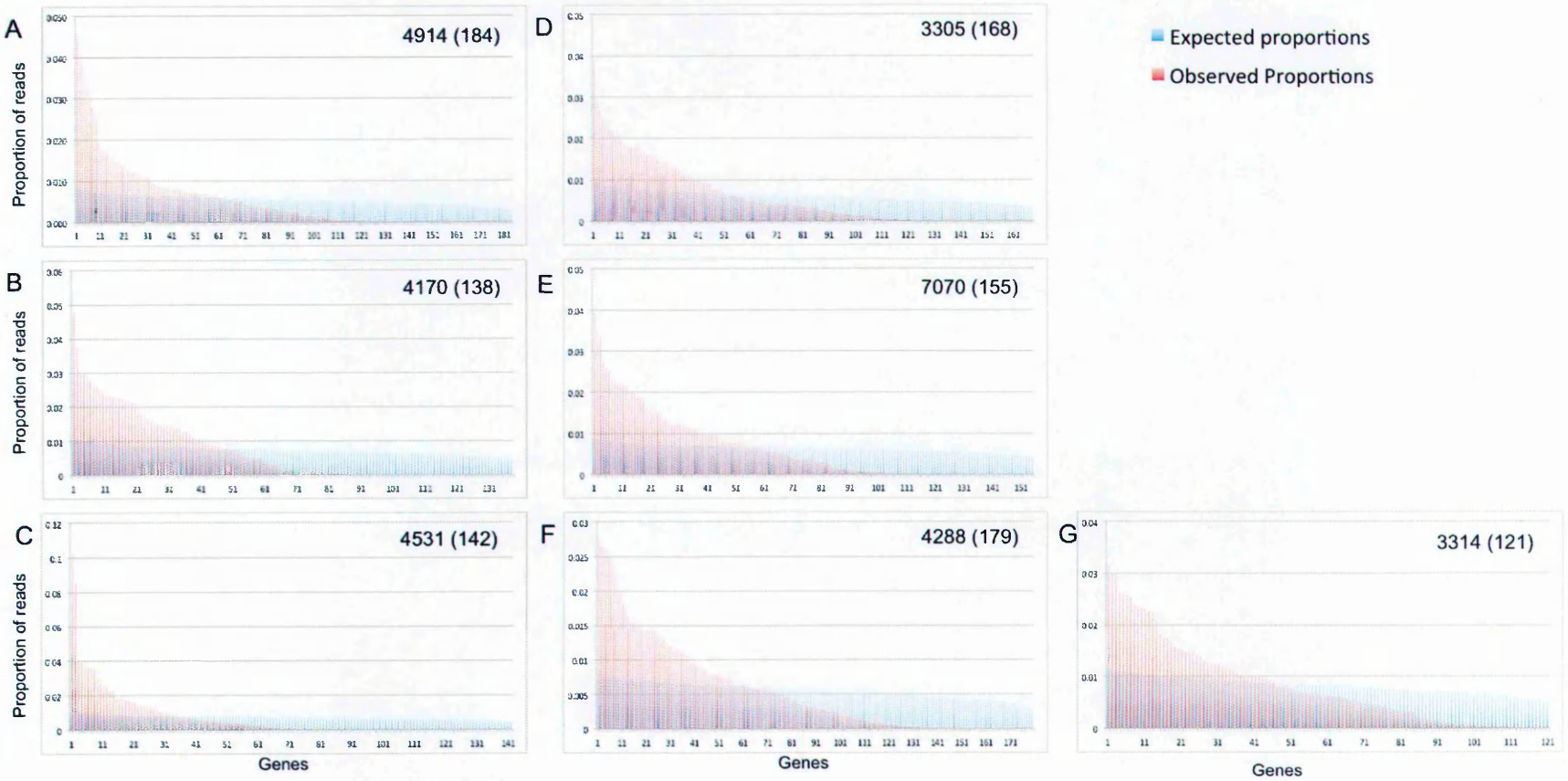
5.4.5. Differential *pir* expression analysis.

A closer look at reads from gDNA data revealed uneven coverage of genomic repertoires in the different parasite isolates and uneven coverage of genes within a genome (Figure 5.7). This was tested by comparing the expected vs the observed coverage of variants per isolate using reads from gDNA amplicons. The expected coverage was calculated in excel using BINOMDIST, using the total number of reads from gDNA as the number of trials, and the probability of getting x number of genes from the annotated *rif/ stevor* genes in a genome. The proportions were then plotted against the observed read counts of each gene from gDNA. There was a difference in the expected and observed frequency distributions for both *rif* (Figure 5.7. I) and *stevor* (Figure 5.7. II), with the primers preferentially amplifying certain *pir* members of the parasite genomes. Genomes from patient isolates have not been manually curated and annotated, and the IT and HB3 genomes are still undergoing editing and correcting, therefore *pir* repertoires for all isolates except 3D7 are incomplete. In order to accurately test for significance between the expected and observed frequency distributions the 3D7 reads data was used. There was a significant difference between the expected and observed coverage for both *rif* and *stevor* repertoires (Pearson's chi squared p=<0.0001) suggesting primer bias. The difference in coverage was required to be taken into account before analysing *rif/stevor* expression across time points. For this reason, the DESeq normalisation approach was used to normalise read data (Anders and Huber, 2010). DESeq was used to normalise for differential sequence coverage by normalising the number of reads from different libraries. After correcting for sequence coverage within an isolate, uneven gene coverage within a genome was accounted for by computing a corrected transcript upregulation by dividing reads from cDNA by reads from gDNA for each gene captured by the primers.

$$\text{Transcript upregulation} = \frac{\text{Normalised cDNA reads}}{\text{Normalised gDNA reads}}$$

Therefore for each isolate both raw read proportions representing raw transcript abundance, and corrected transcript upregulation were computed. Transcript abundance proportions are represented in piecharts, colouring regions of the pie where the transcript took up at least 10% of the total transcription at any cDNA timepoint. Transcript upregulation is represented as heatmaps, with the darker shades of purple indicating increased upregulation. Next to each upregulation heatmap smaller heatmap (orange coloured) was drawn to represent transcript abundance. In this heatmap the reads proportions were grouped into quartiles, with the highest quartile (darker shades of orange) representing high transcript abundance (for example see Figure 5.10 for 3D7 transcript upregulation and abundance).

I


A
4914 (184)
D
3305 (168)
Expected proportions
Observed Proportions
B
4170 (138)
E
7070 (155)
C
4531 (142)
F
4288 (179)
G
3314 (121)
Proportion of reads
Genes

II

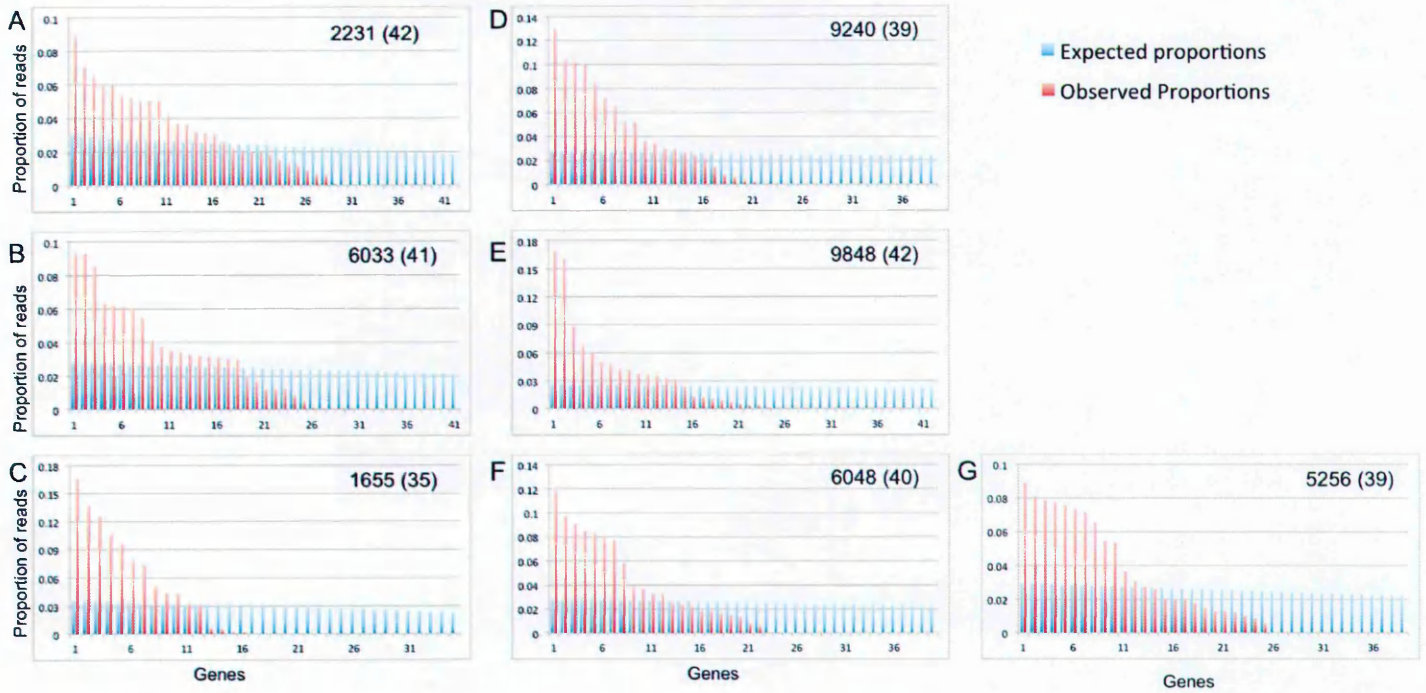


Figure 5.7. *Frequency distribution of expected and observed rif (I) and stevor (II) sequences from gDNA amplicons from 3D7 (A), IT (B), HB3 (C), 9215 (D), 9106 (E), 10668 (F) and SA075 (G) parasites. The number of reads obtained per isolate gDNA amplicons is indicated on each graph the number of rif/ stevor genes annotated in each of the genomes in brackets. Expected distribution was calculated using BINOMDIST in excel.*

5.4.6. A switch in relative proportions of *rifA* to *rifB* transcripts in laboratory isolates and culture-adapted wild isolates.

There are more *rifA* than *rifB* genes in the genome, at about 70% *rifA* to 30% *rifB*, as previously described in chapter 3 (129 *rifA* and 55 *rifB* in 3D7) and this is reflected by the fact that there were generally more *rifA* transcripts than *rifB* transcripts. In laboratory isolates and in three out of the four culture-adapted wild isolates, there was significantly more *rifA* transcripts than *rifB*, than would be expected, at about 90% *rifA* to 10% *rifB* transcripts (Figure 5.8). This difference could not be completely explained by the genomic proportions of *rifA* and *rifB*. In addition, because of better coverage of *rifB* than *rifA* by the primers the proportion *rifA* and *rifB* reads from gDNA amplicons in all isolates was not drastically different (Figure 5.8). This shift in relative transcript proportions in favour of *rifA* was observed in the laboratory isolates, and in 3 of the four culture-adapted wild isolates, with IT and culture-adapted 10668 isolates showing highest rifA transcript proportions (Figure 5.8). In two wild isolates, 9106 and 10668, the culture-adapted lines had markedly preferential *rifA* transcription as compared to ex vivo parasites and HBEC-selected lines (Figure 5.8 E and F), while in the rosetting wild isolate SA075 the bias towards *rifA* transcription was stronger in the rosetting clone than in the non-rosetting fraction. A previous study on VSA expression in laboratory and wild isolates using realtime PCR reported low levels of *rifB* expression (Bachmann et al., 2012). This finding was likely due to poor coverage of *rifB* by the primers used as explained by the authors. In our case primer coverage did not explain the bias. For subsequent *rif* expression profiling the two groups, *rifA* and *rifB* were analysed separately because of this observed bias towards *rifA* transcription.

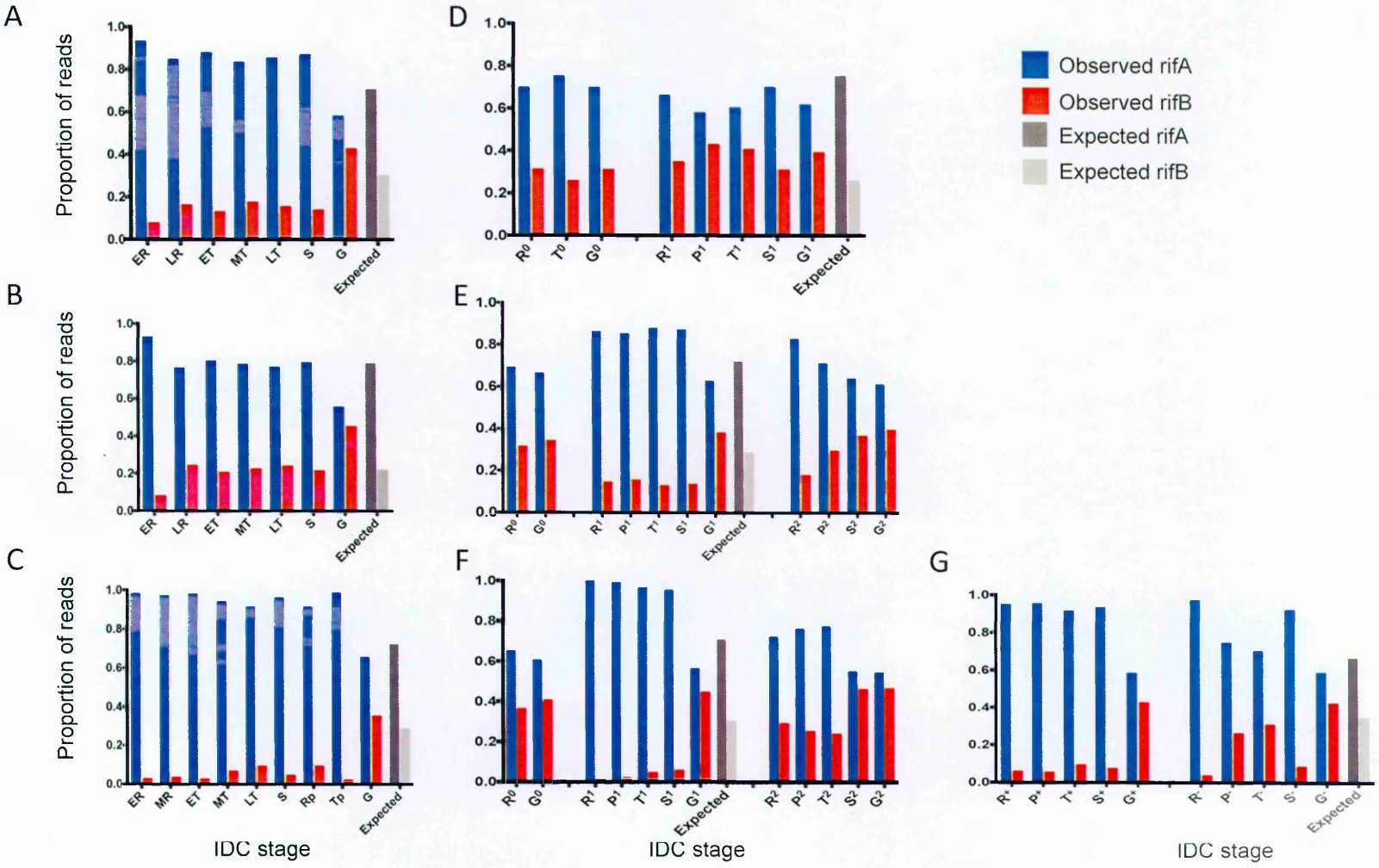

A
B
C
D
E
F
G
Proportion of reads
Observed rifA
Observed rifB
Expected rifA
Expected rifB
ER
LR
ET
MT
LT
S
G
Expected
MR
Rp
Tp
IDC stage

Figure 5.8. *Comparison between rifA and rifB transcript abundance in laboratory and wild isolates. Bar graphs represent proportions of rifA and rifB reads (left Y-axis) number of unique rifA and rifB (right Y-axis) from cDNA and gDNA in 3D7 (A), HB3 (B), IT (C), 9215 (D), 9106 (E), 10668 (F) and SA075 (G) parasites, and selected lines. The X-axis shows the IDC stages, early ring (ER), late ring (LR), early trophozoite (ET), mid trophozoite (MT), late trophozoite (LT) and schizont (S) for the laboratory isolates and rings (R), Late ring to early trophozoites (P), trophozoites (T) and schizonts (S) for the wild isolates. G: proportion of reads from gDNA amplicons. For the Kilifi isolates two to three parasite analysed are shown in the X-axis; Acute parasites (^{0}), culture-adapted parasites (^{1}) and culture-adapted and HBEC-selected (^{2}). For the Kisumu isolate the two lines analysed, the rosette-selected clone (+) and the non-rosetting uncloned fraction (-) are shown in the X-axis. The gray graphs show the expected rifA to rifB ratios from gDNA of the respective isolates.*

5.5. Distinct and dynamic *rif* and *stevor* transcription patterns in laboratory isolates.

Similar to results obtained by cloning and capillary sequencing there was high diversity of *rif* and *stevor* transcripts across the IDC in the laboratory isolates, 3D7 and HB3 having higher diversity than IT. Despite this apparent diversity, more sequencing depth revealed clear stage-specific dominant transcripts, more clearly for *rifB* and *stevor*.

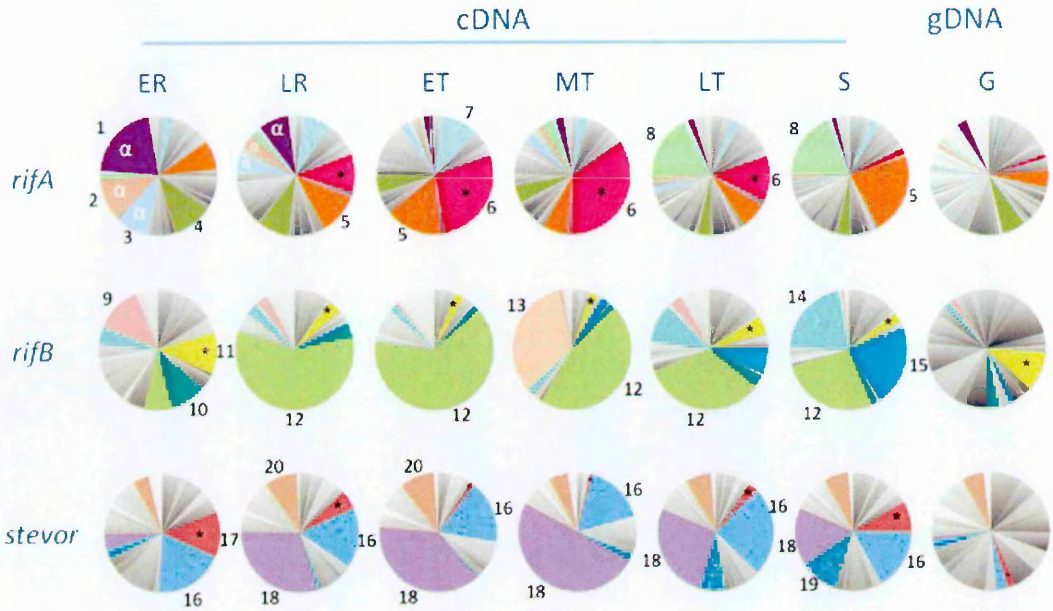
5.5.1. *Pir* expression in 3D7

In 3D7 *rifA* had a high diversity of transcripts while a single transcript dominated *rifB* expression (Figure 5.9). In rings three of the four most abundant *rifA* transcripts were UpsA var-linked-*rif* genes that share a promoter region with UpsA *var,* PF13_0004, PF11_009 and PF11_0520. These have been shown to be upregulated in a 3D7 line that expressed Group A *var* (Wang, 2009) (Figure 5.9). Only one of the *UpsA rif* genes, PF13_0004 was recovered in rings in the capillary data (Figure 4.5 Chapter 4) probably due to reduced coverage from sampling a small number of cloned amplicons. The UpsA *var* linked-*rif* were not expressed in trophozoites and a new set of transcripts PFD1240w, PF10_0006 and a pseudogene, PFD1020c, were most abundant. At schizont stage PFL0025c and PFD1240w were the dominant *rif* transcripts.

rifB had more restricted transcription. In early rings multiple dominant *rifB* transcripts were observed, PF14_0003, PF10_0003, PF10_0006 and a pseudogene PF3D7_0701200. In late rings to trophozoites PF10_0006 dominated *rifB*

transcription. This *rif* has been shown to be expressed in 3D7 early and late IDC stages, sporozoites, mature gametocytes and gametes (Wang et al., 2010, Mwakalinga et al., 2012, Wang et al., 2009). MAL13p1.495 was on in mid trophozoites but was undetectable in late trophozoites. PF10_0006, PF13_0006 and PFI0025c dominated *rifB* expression in schizonts. PF13_0006 and PFI0025c are B1 *rif* that were reported to dominate early and late stage transcription in 3D7 (Wang et al., 2009).

Analysis of *stevor* expression profile showed a difference in expression at early and late stages. There was a switch from the dominant ring transcripts to different ones that stayed on across the later stages of the IDC (Figure 5.8). PFF0850c and PFD0065w transcripts were predominant in rings. There was a switch at the ring-trophozoite transition stage with PF10_0395 being the dominant stevor transcript in late rings and trophozoite stages, but this decreased in schizonts. PFF0850c was maintained at moderate transcription levels all through the IDC. Multiple transcripts were in dominantly transcribed in schizonts with the ring *stevor* PFF0850c and PFD0065w also coming up. PFI0080w was the only stevor transcript that was uniquely transcribed in the late stages. The three dominant *stevor* genes in rings and trophozoites PF10_0395, PFF0850c and PFI0080w had previously been shown to be expressed in 3D7 clones (Niang et al., 2009, Sharp et al., 2006) and gametocytes (Sharp et al., 2006).



Label	*rif/stevor*	New ID	Old ID	*rif*A/B
1	*rif*	PF3D7_1300400	PF13_0004	A^{α}
2	*rif*	PF3D7_1150300	PF11_0520	A^{α}
3	*rif*	PF3D7_1100300	PF11_0009	A^{α}
4	*rif*	PF3D7_0701100	new rifA	A
5	*rif*	PF3D7_0425900	PFD1240w	A
6	*rif ps*	PF3D7_0421500*	PFD1020c*	A*
7	*rif*	PF3D7_0115600	PFA0760w	A
8	*rif*	PF3D7_1200500	PFL0025c	A
9	*rif*	PF3D7_1400300	PF14_0003	B
10	*rif*	PF3D7_1000300	PF10_0003	B
11	*rif ps*	PF3D7_0701200*	new rifB*	B*
12	*rif*	PF3D7_1000600	PF10_0006	B
13	*rif*	PF3D7_1372600	MAL13P1.495	B
14	*rif*	PF3D7_1300600	PF13_0006	B
15	*rif*	PF3D7_0900500	PFI0025c	B
16	stevor	PF3D7_0617600	PFF0850c	na
17	stevor ps	PF3D7_0401500*	PFD0065w*	na
18	stevor	PF3D7_1040200	PF10_0395	na
19	stevor	PF3D7_0901600	PFI0080w	na
20	*stevor*	PF3D7_1479500	PF14_0767	na

Figure 5.9. *(A) 3D7 rif and stevor repertoires and proportion of transcripts in 6 IDC stages, early ring (ER), late ring (LR), early trophozoite (ET), mid trophozoite (MT), late trophozoite (LT) and schizont (S). The pies represent proportions of reads that mapped to rifA (top panel), rifB (middle panel) or stevor (lower panel). Coloured regions of the cDNA pies represent transcripts that accounted for 10% and above of the total expression in at least one time point, while in gDNA pies, coloured regions indicate genes whose read counts represent >10% of the total reads obtained from amplification of matched cDNA. The numbers indicate the gene labels for transcripts that represent 10% and above of the total expression within a time point (see table below). α: UpsA var-linked-rif. *: pseudogene. (B) Table listing transcripts that made up at least 10% of the expression at any stage.*

Due to differential primer coverage of the *rif* repertoire, corrected transcript values were determined by first normalising read counts across time points and gDNA, then dividing normalised read counts from cDNA by normalised gDNA read counts. There was a clear stage-specific upregulation of *rif*, (Figure 5.10). Certain genes clustered together by transcription patterns. For instance *rifs* in cluster 1 and 2 were upregulated through most of the IDC with slight changes while those in cluster 10 were mostly switched off through the IDC.

Clusters 4 and 5 and some members of clusters 6 and 7 were upregulated in the early and late stages. They were switched on in rings, off in mid and late trophozoites, then back on in schizonts. Cluster 4 contained three members of the conserved UpsA-var-linked *rif* group, PF11_0520, PF13_004 and PFD1230c. Previous literature reported this *rif* group to be upregulated in parasites that express the associated *var*. PFD1230c was the most upregulated *rif* in rings with over 100 fold change over gDNA. Interestingly the associated *var* PFD1235w was the second most abundant *var* transcript in 3D7 rings, from the same RNA sample as used to generate this data. This gene *rif* was swiftly switched off in the late ring and stayed at low transcription levels through the rest of the IDC. This profile was strikingly similar to *var* transcription profile.

Cluster 8 and 9 were upregulated in mid and late trophozoites and cluster 3 was kept at low levels throughout the IDC but peaks at the mid trophozoite stage.

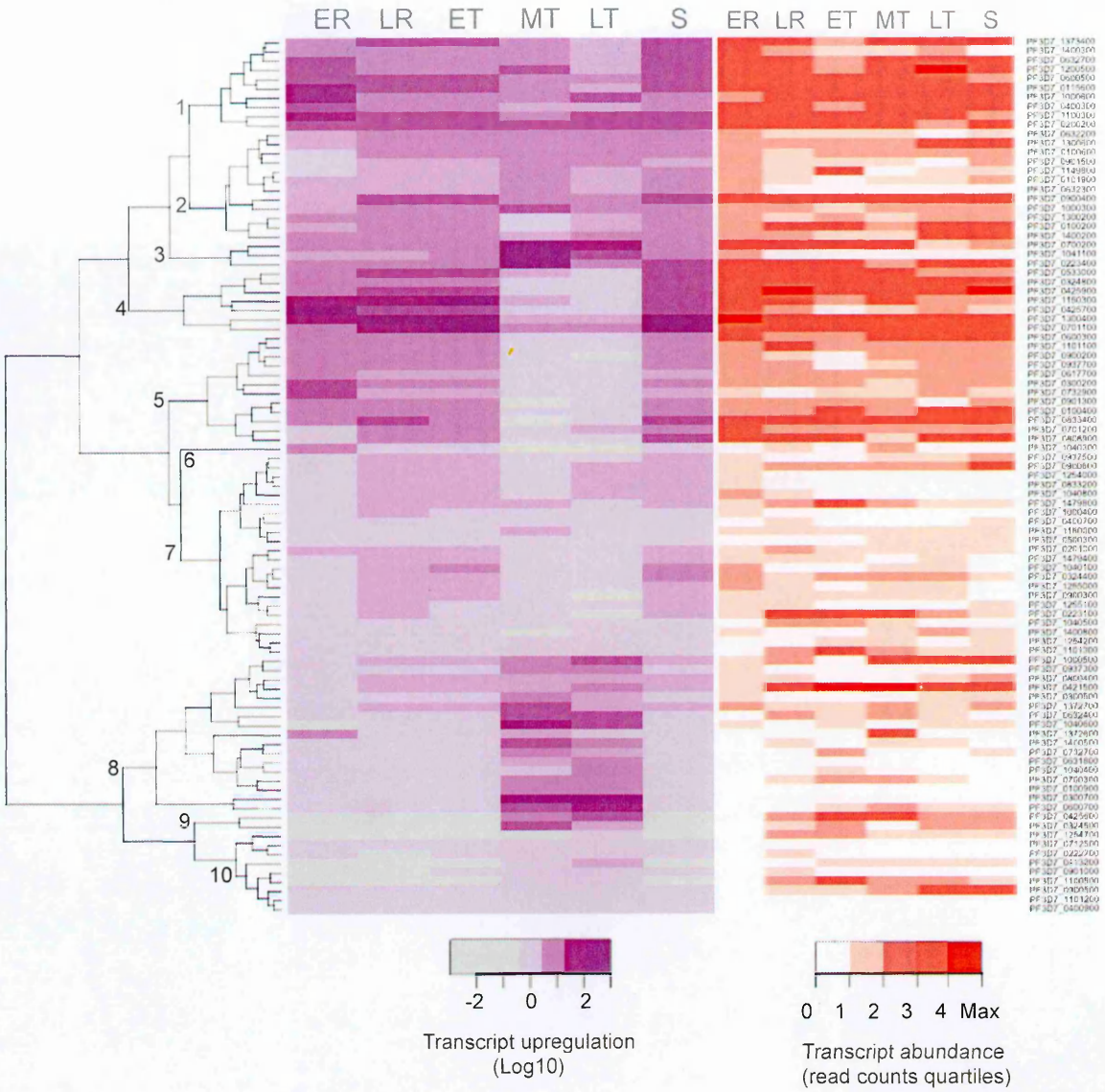


***Figure 5.10**. Heatmaps showing upregulated rif in 3D7 at 6 IDC stages, early ring (ER), late ring (LR), early trophozoite (ET), mid trophozoite (MT), late trophozoite (LT) and schizont (S) (left panel) and transcripts abundance in the same stage (right panel). The panel on the left (purple) shows log10 fold change of the upregulated rif across the IDC. The right panel (orange) represents transcript abundance categorised into quartiles. Dendogram generated by hierarchical clustering in R based on transcript upregulation patterns.*

For *stevor* it was observed that transcript abundance was correlated with transcript upregulation (Figure 5.11). Cluster 1 *stevor* PF10_0395 and PFF0850c, dominant transcripts in mid IDC were the most upregulated transcripts in the same mid IDC stage. Cluster 2 *stevor* were upregulated in early to mid trophozoites and cluster 3 in the late ring early trophozoite transition stage. Cluster 4 and 5 came up in the early and late stages while cluster 6 was stevor that mostly stayed at low transcription levels.

Overall the transcript abundance and upregulation of 3D7 *rif* and *stevor* genes showed good correlation (mean correlation coefficient = 0.82 and 0.9 for *rif* and *stevor* respectively).

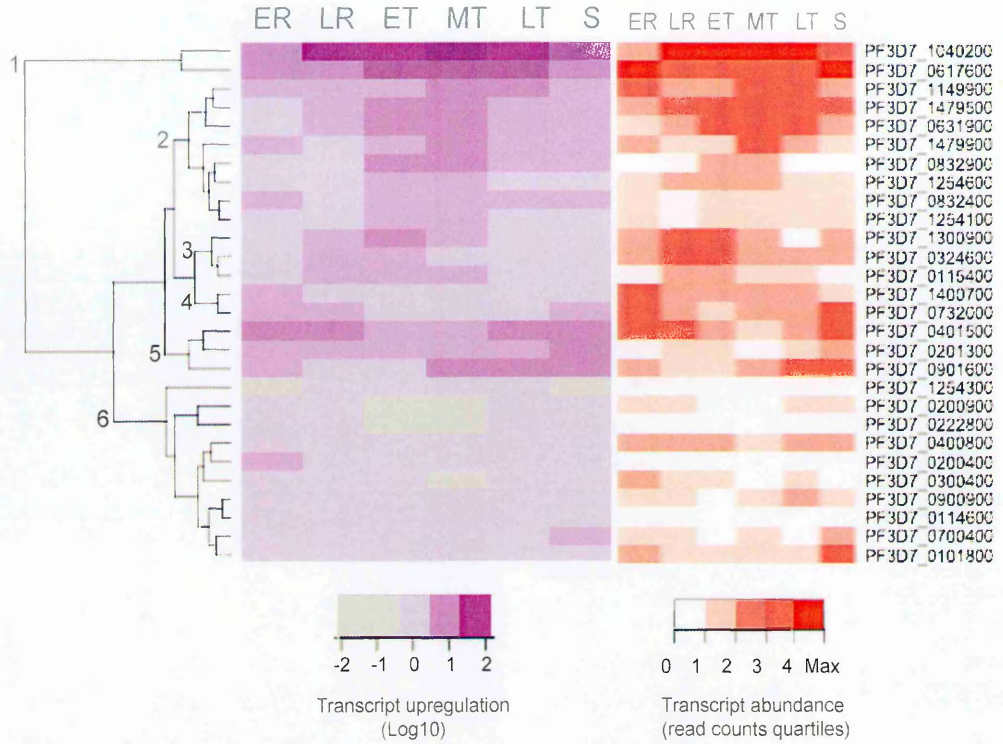


Figure 5.11. *Heatmaps showing stevor transcript upregulation (left panel) and transcript abundance (right panel) in 3D7 at 6 IDC stages, early ring (ER), late ring (LR), early trophozoite (ET), mid trophozoite (MT), late trophozoite (LT) and schizont (S). The panel on the left (purple) shows log10 fold change of the upregulated rif across the IDC. The right panel (orange) represents transcript abundance categorised into quartiles. Dendogram generated by hierarchical clustering in R based on transcript upregulation patterns.*

5.5.2. *Pir* expression in IT

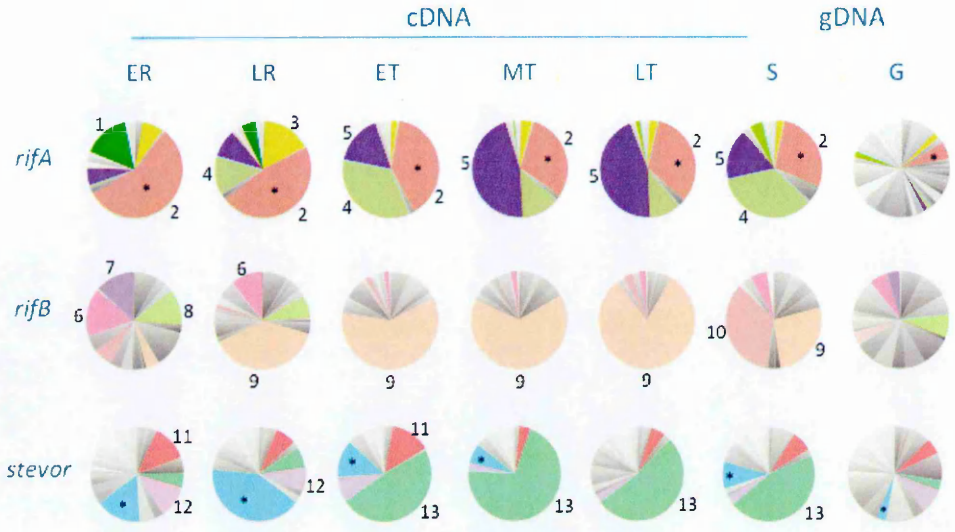
rif and stevor transcription profiles were explored in two IT samples, an original aliquot grown from the IT obtained from Sanger (IT original) and another that was kept in continuous culture in our laboratory for over 40 cycles (IT passaged). The *rifA* to *rifB* transcript proportions were more skewed to *rifA* in IT as compared to other laboratory isolates analysed (Figure 5.8. C). Despite the high proportion of *rifA* transcripts, the number of unique *rifA* was much smaller than what was observed in the other 2 laboratory isolates, with only 5 dominant transcripts across the IDC (Figure 5.12. A). This restricted diversity of transcripts was previously observed in the capillary sequencing data (Chapter 4 figure 4.6). A single *rifA* pseudogenic transcript PFIT_0536700 dominated the first half of the IDC (ER-ET) and was maintained at high levels in the rest of the IDC. This *rif* is related to the conserved 3D7 pseudogene PFE1635w that was expressed in low levels in the early stages in 3D7 (Appendix table 9.8). In IT and 3D7 genomes this *rif* (PFITbin2300), that was previously observed to be highly transcribed in rings in the capillary data, was also highly expressed in rings in the 454 data. Two other *rifA* genes, PFIT_0835500 and PFIT_0900150, were dominantly transcribed in trophozoite and schizont stages as previously observed in our capillary data (Figure 4.6 A Chapter 4). There was no significant change in *rifA* transcription in IT passaged in culture (Figure 5.12. B)

There was a much more restricted diversity of *rifB* in mid to late stages (Figure 5.12. A). Three dominant transcripts were observed in early rings, and a single dominant transcript PFIT_0733200 came up in late rings and dominated transcription in the later stages. This *rif* is 99% identical to the 3D7 MAL13P1.495, which came up in mid trophozoites (Figure 5.9 sequence "13"). There was very little difference in *rifB*

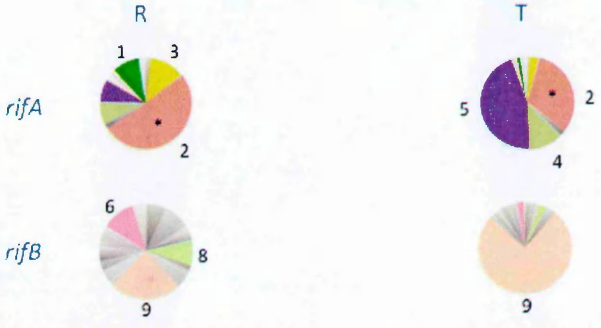
transcription between IT original and IT passaged. The transcript PFIT_bin10000 was dominantly transcribed in original IT ring stage but not recovered in the passage IT line (Figure 5.12. B).

Stevor showed a similar transcription pattern as *rifB,* with multiple transcripts in rings and a switch to a single dominant transcript in the later stages. In rings two *stevor* PFIT_0700100 and PFIT_bin00600, and a pseudogene PFIT_0400200 were dominantly transcribed. The pseudogene PFIT_0400200 is 95% identical to the 3D7 pseudogene PFD0065w. This was the dominant ring transcript and was maintained at a low level through out the rest of the IDC. In the late ring stage there was a switch to a new dominant transcript PFIT_bin03800 that was maintained at high levels through the later stages of the IDC (Figure 5.12. B). In late rings to schizonts PFIT_bin03800 was the most abundant *stevor* transcript.

A



B



C

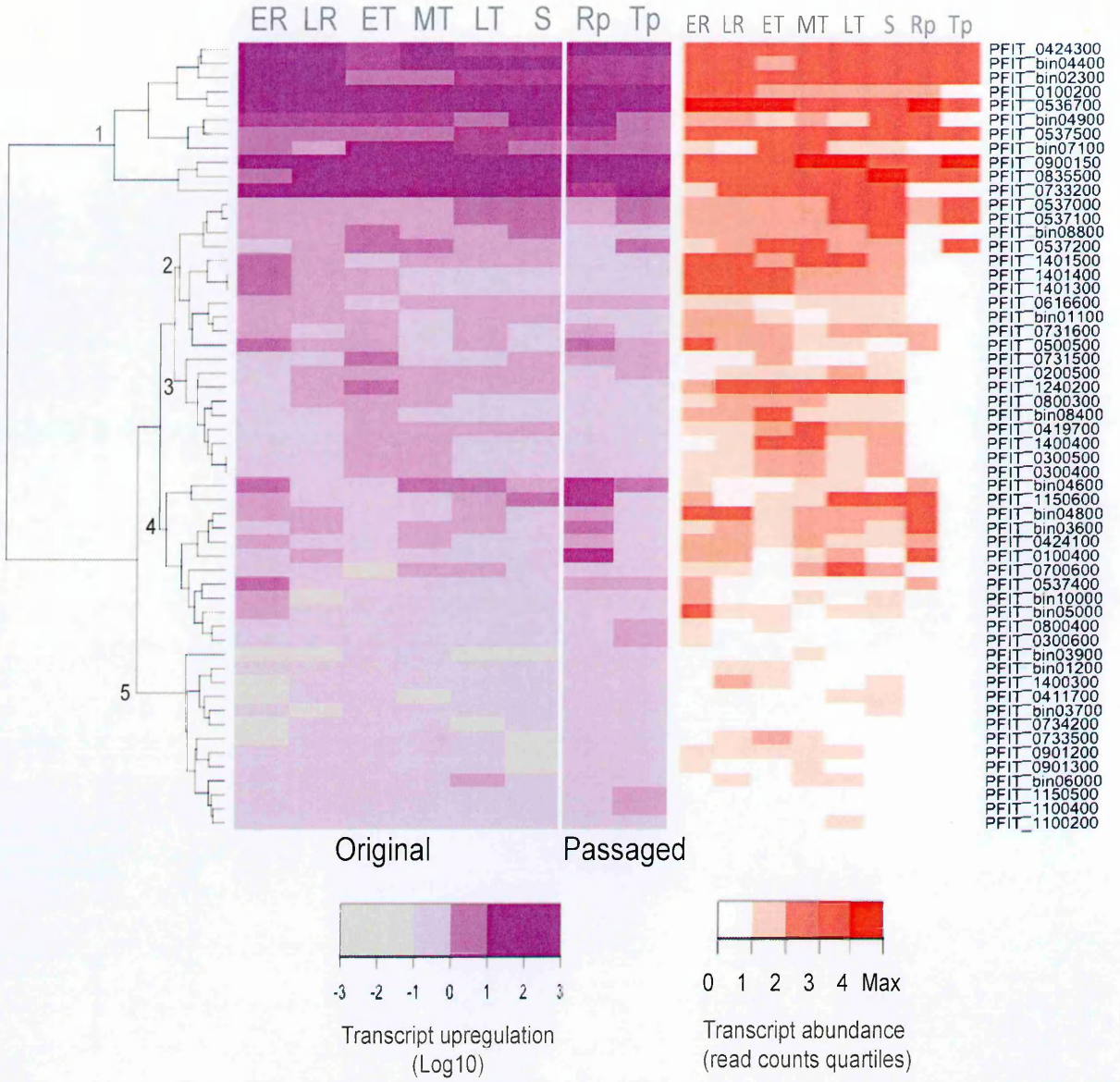
Label	*rif/stevor*	ID	*rif*A/B	3D7 BLAST hit	3D7 old gene ID	Identities
1	*rif*	PFIT_bin02300[b]	A	PF3D7_1372700	MAL13P1.500	681/889 (77%)
2	*rif* (c)	PFIT_0536700*	A*	PF3D7_0533000*	PFE1635w*	813/853 (95%)
3	*rif*	PFIT_0424300	A	PF3D7_0500400	PFE0020c	511/618 (82%)
4	*rif*	PFIT_0835500[a]	A	PF3D7_0901400	PFI0070w	809/1061 (76%)
5	*rif*	PFIT_0900150	A	PF3D7_0833400	MAL7P1.213	761/971 (78%)
6	*rif*	PFIT_bin04800	B	PF3D7_0201000	PFB0055c	761/952 (80%)
7	*rif*	PFIT_bin10000	B	PF3D7_1000300	PF10_0003	874/1064 (82%)
8	*rif*	PFIT_0424100	B	PF3D7_0324500	PFC1100w	789/886 (89%)
9	*rif* (c)	PFIT_0733200	B	PF3D7_1372600	MAL13P1.495	817/819 (99%)
10	*stevor*	PFIT_0700100	na	PF3D7_0832400	MAL7P1.223	699/865 (80%)
11	*stevor*	PFIT_bin00600	na	PF3D7_1479500	PF14_0767	747/835 (89%)
12	*stevor* (c)	PFIT_0400200*	na	PF3D7_0401500*	PFD0065w*	863/864 (99%)
13	*stevor*	PFIT_bin03800	na	PF3D7_0732000	PF07_0130	705/838 (84%)

Figure 5.12. *rif and stevor repertoires and proportion of transcripts in (A) IT (original) at 6 IDC stages, early ring (ER), late ring (LR), early trophozoite (ET), mid trophozoite (MT), late trophozoite (LT) and schizont (S) and (B) IT(passaged) at 2 IDC stages, mixed rings (R) and mixed trophozoites (T). The pies represent proportions of reads that mapped to rifA (first panel), rifB (second panel) or stevor (third panel). Coloured regions of the cDNA pies represent transcripts that accounted for 10% and above of the total expression in at least one time point, while in gDNA pies, coloured regions indicate genes whose read counts represent >10% of the total reads obtained from amplification of matched cDNA. The numbers indicate the gene labels for transcripts that represent 10% and above of the total expression within a time point (table C). *: pseudogene. (C) Table showing most abundant rif/stevor transcripts in IT. * pseudogene.* [a] *rif1, PaloAlto (PAR+/PAR-) (Kyes et al., 1999),* [b] *rif3, PaloAlto (PAR-) (Kyes et al., 1999).*

In IT *rif* transcript upregulation was weakly correlated with transcript abundance (mean correlation coefficient of and 0.45). Cluster 1, which contains the 5 dominant *rifA* and 1 dominant *rifB* transcript, is upregulated in most IDC stages (Figure 5.13. A). The two dominant ring transcripts PFIT_bin02300 and PFIT_0536700 were among the most highly upregulated in rings and group together within cluster 1. Cluster 2, which is purely made up of *rifA*, appears to have no clear upregulation pattern with some members coming up in the early stages, some in the middle stage and others in the late stage. Cluster 3 appears to be upregulated in the mid stages. Cluster 4, which mostly contained *rifB*, is upregulated mainly in the early stages with the majority showing two peaks of upregulation in the early and middle stages. This cluster contains all the dominant *rifB* transcripts, suggesting that the most abundant *rifB* in IT have similar transcription profiles. Members of cluster 5 do not appear to be upregulated in any stage.

Stevor transcript abundance and upregulation correlated well in IT (mean correlation coefficient of 0.95) (Figure 5.13. B). The most abundant transcripts fell into cluster 1 with the exception of PFIT_bin00600, which fell in cluster 2. The dominant ring transcript PFIT_0700100 was upregulated in the early to mid stages (rings to trophozoites) while the dominant mid and late stage transcript PFIT_bin03800 was upregulated in mid to late stages (early trophozoites to schizonts). Cluster 4 was upregulated mainly in the late stages while cluster 5 was slightly up in early and late stages, patterns similar to their peak transcription time points from the transcript abundance analysis (Figure 5.12).

A


ER LR ET MT LT S Rp Tp
ER LR ET MT LT S Rp Tp
1
2
3
4
5
PFIT_0424300
PFIT_bin04400
PFIT_bin02300
PFIT_0100200
PFIT_0536700
PFIT_bin04900
PFIT_0537500
PFIT_bin07100
PFIT_0900150
PFIT_0835500
PFIT_0733200
PFIT_0537000
PFIT_0537100
PFIT_bin08800
PFIT_0537200
PFIT_1401500
PFIT_1401400
PFIT_1401300
PFIT_0616600
PFIT_bin01100
PFIT_0731600
PFIT_0500500
PFIT_0731500
PFIT_0200500
PFIT_1240200
PFIT_0800300
PFIT_bin08400
PFIT_0419700
PFIT_1400400
PFIT_0300500
PFIT_0300400
PFIT_bin04600
PFIT_1150600
PFIT_bin04800
PFIT_bin03600
PFIT_0424100
PFIT_0100400
PFIT_0700600
PFIT_0537400
PFIT_bin10000
PFIT_bin05000
PFIT_0800400
PFIT_0300600
PFIT_bin03900
PFIT_bin01200
PFIT_1400300
PFIT_0411700
PFIT_bin03700
PFIT_0734200
PFIT_0733500
PFIT_0901200
PFIT_0901300
PFIT_bin06000
PFIT_1150500
PFIT_1100400
PFIT_1100200
Original
Passaged
-3 -2 -1 0 1 2 3
Transcript upregulation
(Log10)
0 1 2 3 4 Max
Transcript abundance
(read counts quartiles)

B

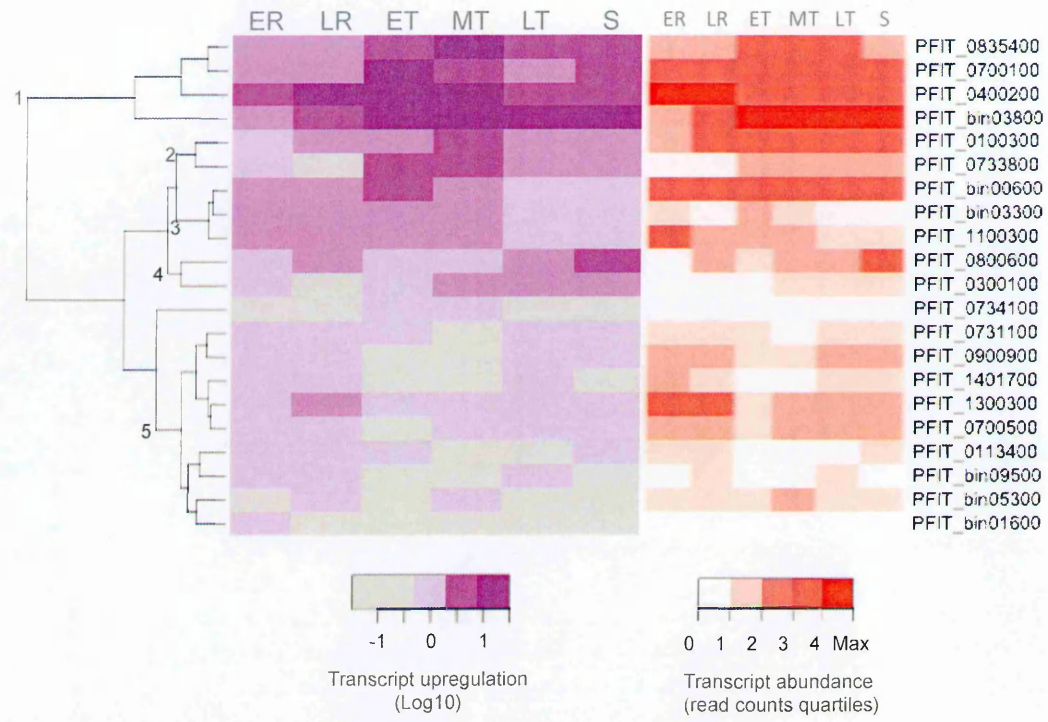


***Figure 5.13**. Heatmaps showing rif (A) and stevor (B) transcript upregulation on the left panel and transcript abundance on the right panel in IT with the top heatmap showing rif patterns in IT (original) at 6 IDC stages, early ring (ER), late ring (LR), early trophozoite (ET), mid trophozoite (MT), late trophozoite (LT) and schizont (S), and I T(passaged) at ring (Rp) and trophozoite (Tp) stages. The two heatmaps on the left (purple) shows log10 fold change of upregulated rif across the IDC while the right panel (orange) represents transcript abundance categorised into quartiles. The lower heatmaps show stevor patterns in IT (original) across 6 IDC stages. The dendogram was generated by hierarchical clustering in R based on transcript upregulation patterns.*

5.5.3. *Pir* expression in HB3

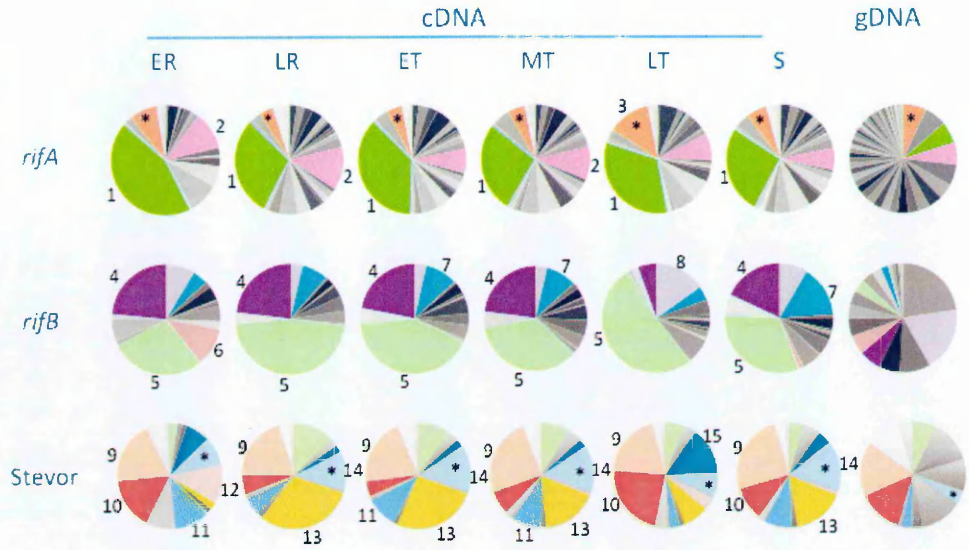
For HB3 there was very little change in *rif* transcription across the IDC (Figure 5.14). The three most abundant *rifA* transcripts, PfHB3_BroadWash530300, PfHB3_BroadWash520600 and PfHB3_BroadWash536100 were maintained across the IDC with PfHB3_BroadWash530300 being the dominant *rifA*. The dominant *rifA* was 94% identical to the IT *rif* PFIT_1401300 whose transcription maxima was in early rings and 92% identical to the 3D7 PFI0020w also maximally transcribed in rings. Unlike in IT and 3D7 in HB3 this *rif* is maintained at high transcription levels through out the IDC. Another *rifA* and *rifA* pseudogene were also dominantly expressed across the IDC.

PfHB3_BroadWash530200 and PfHB3_BroadWash551000 were the most abundant *rifB* transcripts with PfHB3_BroadWash530200 predominating transcription. PfHB3_BroadWash262300 came up in the mid to late stages while PfHB3_BroadWash161200 peaked in late trophozoites.

Similarly there was little change in *stevor* transcription across the IDC. PfHB3_BroadWash161300 and PfHB3_BroadWash532500 were the most abundant transcripts in early rings with PfHB3_BroadWash161300 staying up in the rest of the IDC. Two other dominant ring transcripts were observed (Figure 5.14). PfHB3_BroadWash521400 dominated transcription in the late rings to mid trophozoites, and schizonts. A schizont *stevor* transcript PfHB3_BroadWash544800 belongs to a conserved *stevor* pseudogene group that includes the 3D7 PFD0065w,

transcribed in rings and schizonts (Figure 5.9 sequence "17") and expressed in all the laboratory isolates (Figure 5.9 and Figure 5.12).

A



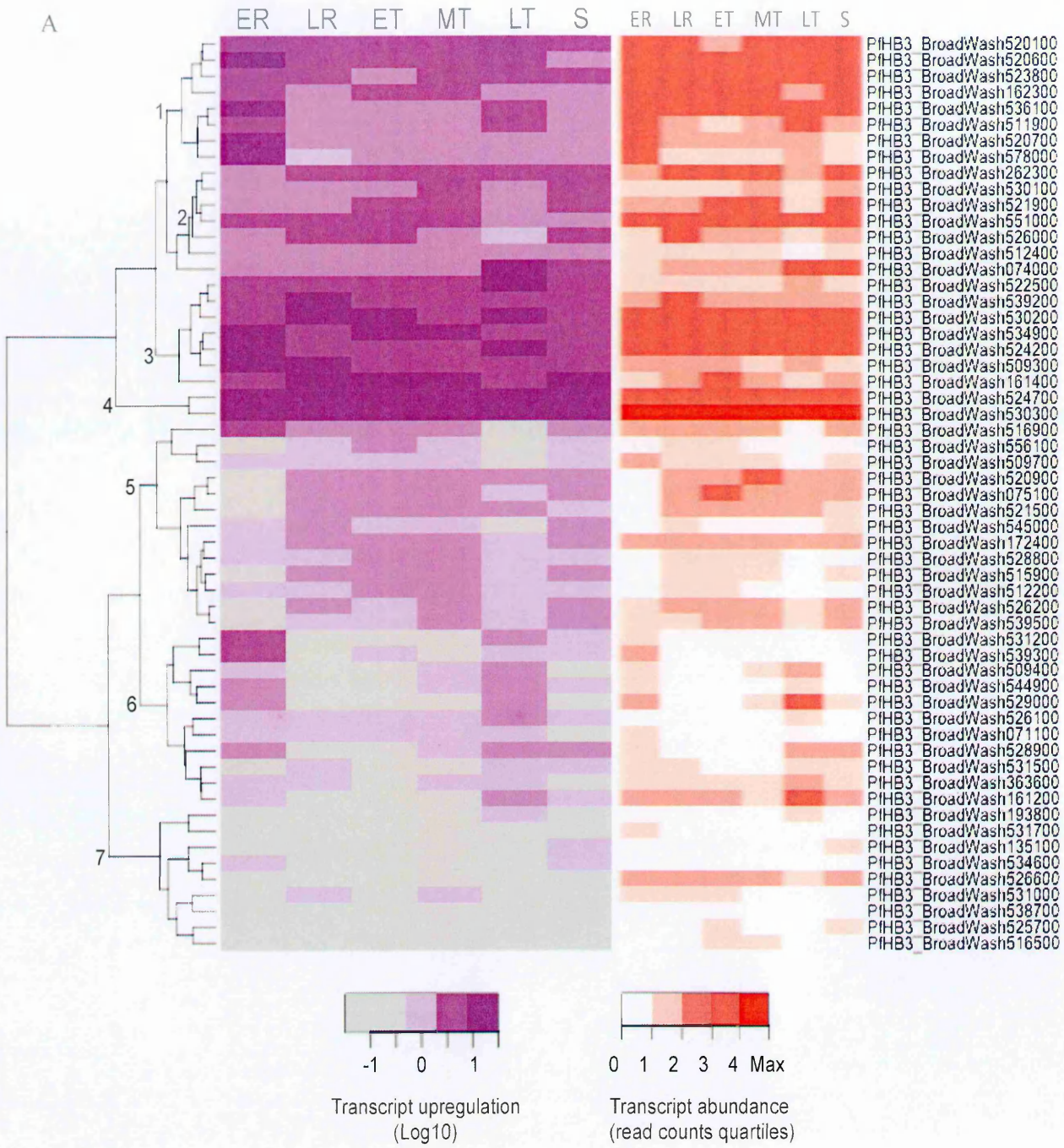
B

Label	*pir*	ID	*rifA/B*	3D7/ IT BLAST hit	Identities
1	*rif*	PfHB3_BroadWash530300	A	PFIT_1401300	975/1035 (94%)
2	*rif*	PfHB3_BroadWash520600	A	PF3D7_0324400 (PFC1095w)	858/1059 (81%)
3	*rif* (c)	PfHB3_BroadWash536100*	A*	PF3D7_0533000 (PFE1635w)*	633/639 (99%)
4	*rif*	PfHB3_BroadWash551000	B	PF3D7_0425900 (PFD1240w)	497/596 (83%)
5	*rif*	PfHB3_BroadWash530200	B	PFIT_bin01700	807/957 (84%)
6	*rif*	PfHB3_BroadWash529000	B	PFIT_0100400	885/934 (94%)
7	*rif* (c)	PfHB3_BroadWash262300	B	PF3D7_1000600 (PF10_0006)	1061/1086 (97%)
8	*rif*	PfHB3_BroadWash161200	B	PF3D7_0701200*	694/887 (78%)
9	*stevor*	PfHB3_BroadWash161300	na	PFIT_0700500	734/800 (91%)
10	*stevor*	PfHB3_BroadWash532500	na	PF3D7_1254100 (PF07_0130)	798/898 (88%)
11	*stevor*	PfHB3_BroadWash528600	na	PFIT_bin09500*	551/558 (98%)
12	*stevor*	PfHB3_BroadWash162000	na	PF3D7_0300400	809/893 (90%)
13	*stevor*	PfHB3_BroadWash521400	na	PF3D7_125410 (PFL2610w)	767/892 (85%)
14	*stevor* (c)	PfHB3_BroadWash544800*	na	PF3D7_0401500 (PFD0065w)*	451/458 (98%)
15	*stevor*	PfHB3_BroadWash538600	na	PF3D7_1149900 (PF11_0516)	816/933 (87%)

Figure 5.14. *HB3 rif and stevor repertoires and proportion of transcripts in 6 IDC stages, early ring (ER), late ring (LR), early trophozoite (ET), mid trophozoite (MT), late trophozoite (LT) and schizont (S). The pies represent proportions of reads that mapped to rifA (top panel), rifB (middle panel) or stevor (lower panel). Coloured regions of the cDNA pies represent transcripts that accounted for 10% and above of the total expression in at least one time point, while in gDNA pies, coloured regions indicate genes whose read counts represent >10% of the total reads obtained from amplification of matched cDNA (see table below). The numbers indicate the gene labels for transcripts that represent 10% and above of the total expression within a time point.*

Rif transcript upregulation and transcript abundance were weakly correlated in HB3 (mean correlation coefficient = 0.52). There was a temporal pattern of transcript upregulation (Figure 5.15. A). Cluster 1, which was made up purely of *rifA*, was mostly upregulated in the early rings, with a sub cluster staying on in the middle and late stages. Cluster 3 and 4, which contained the dominant *rifA* and *rifB* transcripts respectively, were upregulated through most of the IDC. Cluster 5 was upregulated in late rings to mid trophozoites, switched off in late trophozoites and back on in schizonts. Cluster 6 was upregulated in the early and late stages. Cluster 7 was mostly off all through the IDC (Figure 5.15).

Similarly, the correlation between *stevor* transcript upregulation and transcript abundance was weak (mean correlation coefficient = 0.33). Cluster 1 was upregulated in early and late stages. Cluster 2 was up in early and middle stages, with some members staying on in late stages. Cluster 3 was predominantly upregulated in late trophozoites. Cluster 4 and 5 were mostly turned off (Figure 5.15. B).


A
ER LR ET MT LT S
ER LR ET MT LT S
1
2
3
4
5
6
7
PfHB3_BroadWash520100
PfHB3_BroadWash520600
PfHB3_BroadWash523800
PfHB3_BroadWash162300
PfHB3_BroadWash536100
PfHB3_BroadWash511900
PfHB3_BroadWash520700
PfHB3_BroadWash578000
PfHB3_BroadWash262300
PfHB3_BroadWash530100
PfHB3_BroadWash521900
PfHB3_BroadWash551000
PfHB3_BroadWash526000
PfHB3_BroadWash512400
PfHB3_BroadWash074000
PfHB3_BroadWash522500
PfHB3_BroadWash539200
PfHB3_BroadWash530200
PfHB3_BroadWash534900
PfHB3_BroadWash524200
PfHB3_BroadWash509300
PfHB3_BroadWash161400
PfHB3_BroadWash524700
PfHB3_BroadWash530300
PfHB3_BroadWash516900
PfHB3_BroadWash556100
PfHB3_BroadWash509700
PfHB3_BroadWash520900
PfHB3_BroadWash075100
PfHB3_BroadWash521500
PfHB3_BroadWash545000
PfHB3_BroadWash172400
PfHB3_BroadWash528800
PfHB3_BroadWash515900
PfHB3_BroadWash512200
PfHB3_BroadWash526200
PfHB3_BroadWash539500
PfHB3_BroadWash531200
PfHB3_BroadWash539300
PfHB3_BroadWash509400
PfHB3_BroadWash544900
PfHB3_BroadWash529000
PfHB3_BroadWash526100
PfHB3_BroadWash071100
PfHB3_BroadWash528900
PfHB3_BroadWash531500
PfHB3_BroadWash363600
PfHB3_BroadWash161200
PfHB3_BroadWash193800
PfHB3_BroadWash531700
PfHB3_BroadWash135100
PfHB3_BroadWash534600
PfHB3_BroadWash526600
PfHB3_BroadWash531000
PfHB3_BroadWash538700
PfHB3_BroadWash525700
PfHB3_BroadWash516500
-1 0 1
Transcript upregulation
(Log10)
0 1 2 3 4 Max
Transcript abundance
(read counts quartiles)

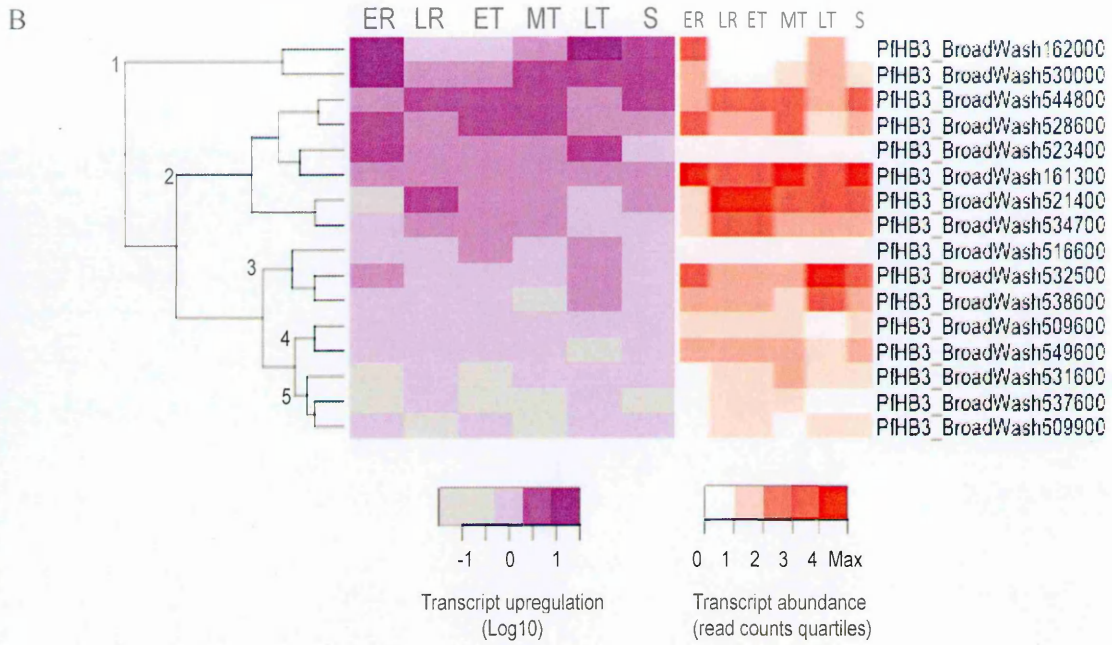


***Figure 5.15**. Heatmaps showing rif (A) and stevor (B) transcript upregulation (purple) and transcript abundance (orange) in HB3 at 6 IDC stages, early ring (ER), late ring (LR), early trophozoite (ET), mid trophozoite (MT), late trophozoite (LT) and schizont (S). The left panel (purple) indicates log10 fold change transcript upregulation across the IDC. The right panel (orange) represents transcript abundance categorised into quartiles. The dendogram was generated by hierarchical clustering in R based on transcript upregulation patterns.*

5.6. *rif* and *stevor* expression patterns in wild isolates under different selection pressures.

To understand whether selection pressure has any effect on *rif* and *stevor* gene expression, analysis of transcription patterns was carried out in wild isolates from Kilifi and Kisumu that had undergone different selection pressure, within host selection pressure, culture-adaptation, and selection for binding to human brain endothelial cell line (HBEC) or rosetting. *Rif* and *stevor* profiling was carried out at different time points in two to three different lines per isolate; acute (fresh-from-the-arm) rings (R) and ex vivo trophozoite (T), rings (R), late rings to early pigmented trophozoites (LR/ET or P), mid and late trophozoites (T) and schizont (S) in culture-adapted lines and selected lines.

5.6.1. *Rif* and *stevor* transcription in acute and culture-adapted 9215

Rif and *stevor* transcription was analysed in 9215 acute and culture-adapted parasites. This parasite originated from a patient with severe malaria with respiratory distress. Although the same sets of transcripts were observed in both the acute and culture-adapted lines there was a change in the relative proportions (Figure 5.16). Multiple *rifA* transcripts were recovered in the acute rings with no clear dominants. Two of the ring transcripts PF8443_8_4_U539500 and PF8443_8_4_U509100 dominated transcription in the acute trophozoites. After culture adaptation one of the acute transcripts PF8443_8_4_U539500 dominated ring *rifA* transcription and was maintained all through the IDC while two other minor *rifA* transcripts PF8443_8_4_U522000 and PF8443_8_4_U509100 became more prominent. The same dominant *rifA* transcripts in acute trophozoites were observed in cultured

trophozoites. A conserved *rifA* pseudogene PF8443_8_4_U512100 (PFE1635w in 3D7) came up in the culture-adapted line and was expressed in all time points.

For *rifB* the dominant acute ring transcript PF8443_8_4_U539500 was also dominant in culture. A minor transcript in the acute parasite PF8443_8_4_U512100 became dominant in culture peaking at rings and schizonts. The dominant *rifB* in acute trophozoites PF8443_8_4_U537200 was turned off in culture.

Multiple dominant *stevor* transcripts were observed in acute rings. One of these, PF8443_8_4_U026200, the conserved *stevor* pseudogene (PFD0065w in 3D7) became the dominant transcript in the ex vivo trophozoites, and in the cultured rings and early trophozoites. One of the other dominant *stevor* transcripts in acute rings, PF8443_8_4_U546300, was dominantly transcribed in the culture-adapted line in the late ring to early trophozoite stage. PF8443_8_4_U533600 that was a minor transcript in acute parasites was predominantly transcribed in cultured trophozoites and schizonts (Figure 5.16).

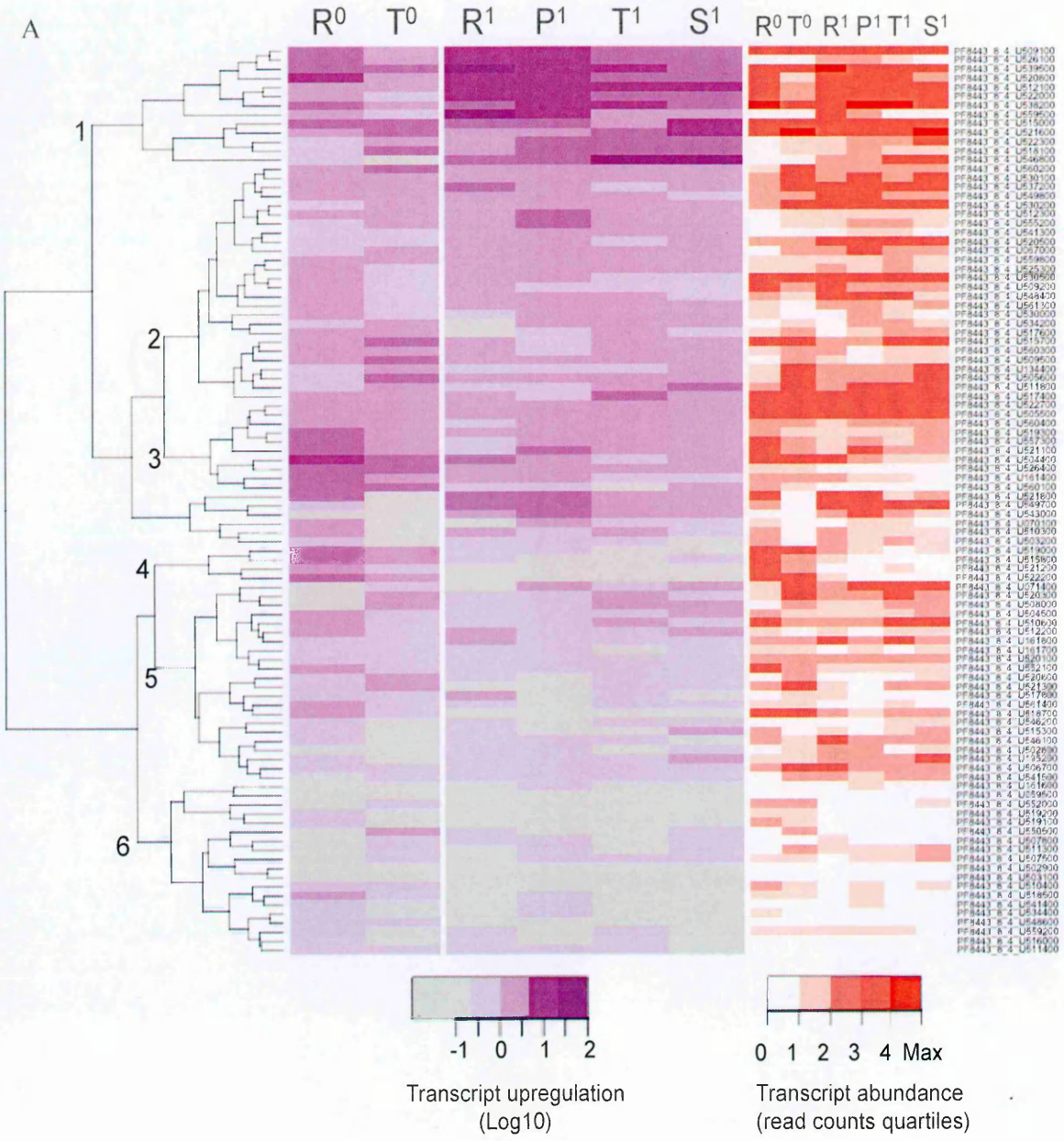
9215 RD	Ring cDNA			LR / ET cDNA			Trophozoite cDNA			Schizont cDNA			Genomic DNA		
	rifA	*rifB*	*stev*	*rifA*	*rifB*	*stev*	*rifA*	*rifB*	*stev*	*rifA*	*rifB*	*stev*	*rifA*	*rifB*	*stev*
															
															

colour	*rif/stevor*	ID	A/B	Top BLAST hit	Identities	2nd BLAST hit #	Identities
	rif	PF8443_8_4_U539500	A	PF8443_8_7_U514500	658/715 (92%)	PF3D7_1254800 PFL2645c)	716/853 (83%)
	rif	PF8443_8_4_U521600	A	PF3D7_0200200	708/827 (85%)	PF8443_8_6_U525100	697/834 (83%)
	rif	PF8443_8_4_U522000	A	PF8443_8_3_U508000	472/569 (82%)	PF3D7_0901300 (PFI0065w)	305/337 (90%)
	rif	PF8443_8_4_U509100	A	PFSAO75_PN54	525/603 (87%)	PF3D7_0600700 (PFF0035c)	401/470 (85%)
	rif (c)	PF8443_8_4_U512100*	A*	PF8443_8_6_U537300	602/615 (97%)	PFIT_0536700 (PFE1635w)*	585/589 (99%)
	rif	PF8443_8_4_U505500	B	RIF_K355 (10739)	646/730 (88%)	PF3D7_0222500 (PFB1005w)	749/892 (83%)
	rif	PF8443_8_4_U538200	B	PF8443_8_6_U516500	917/918 (99%)	PfHB3_BroadWash551000	917/918 (99%)
	rif	PF8443_8_4_U537200	B	PFIT_0500300	803/919 (87%)	PF8443_8_6_U546400	793/917 (86%)
	stevor (c)	PF8443_8_4_U026200*	na	PF8443_8_1_U501600*	572/572 (100%)	PF3D7_0401500 (PFD0065w)*	344/344 (100%)
	stevor	PF8443_8_4_U533600	na	PF8443_8_5_U157400	840/857 (98%)	PFIT_0400200	562/611 (91%)
	stevor	PF8443_8_4_U516100	na	PfHB3_BroadWash531600	842/914 (92%)	PF3D7_1479500 (PF14_0767)	836/916 (91%
	stevor	PF8443_8_4_U265000	na	PFSAO75_258600	628/628 (100%)	PF3D7_1400700 (PF14_0007)	585/625 (93%)
	stevor	PF8443_8_4_U503400	na	PFIT_1401700	798/883 (90%)	PF8443_8_7_U557900	745/833 (89%)

Figure 5.16. *rif and stevor genomic repertoires and transcription profiling in 9215 at acute rings and ex vivo trophozoites (top panel) and after culture adaptation at 4 time points. The pies represent proportions of reads that mapped to rifA (top panel), rifB (middle panel) or stevor (lower panel). Coloured regions of the cDNA pies represent transcripts that accounted for 10% and above of the total expression in at least one time point, while in gDNA pies, coloured regions indicate genes whose read counts represent >10% of the total reads obtained from amplification of matched cDNA (see table).*

For some *rifs* there was a clear upregulation pattern, especially in the cultured parasites, but for most there was little difference in upregulation between acute and cultured parasites (Figure 5.17. A). The *rif* transcripts upregulation in the cultured parasites belonged to cluster 1. This cluster contained the dominantly transcribed *rifs*, which were more clearly seen in the cultured lines than in the acute/ ex vivo parasites. Cluster 3 and 4 were mostly upregulated in acute rings.

One *stevor* was markedly upregulated in cultured parasites. PF8443_8_4_U533600 was significantly upregulated in mid and late stage parasites in the culture-adapted line (Figure 5.17. B). PF8443_8_4_U026200 was upregulated in acute trophozoites but in culture this *stevor* was most upregulated in rings (Figure 5.17. B). While most *stevors* stayed constant in acute and cultured parasites some were marginally upregulated in acute rings and cultured trophozoites and schizonts.


A
R0 T0 R1 P1 T1 S1
R0 T0 R1 P1 T1 S1
1
2
3
4
5
6
-1 0 1 2
Transcript upregulation
(Log10)
0 1 2 3 4 Max
Transcript abundance
(read counts quartiles)

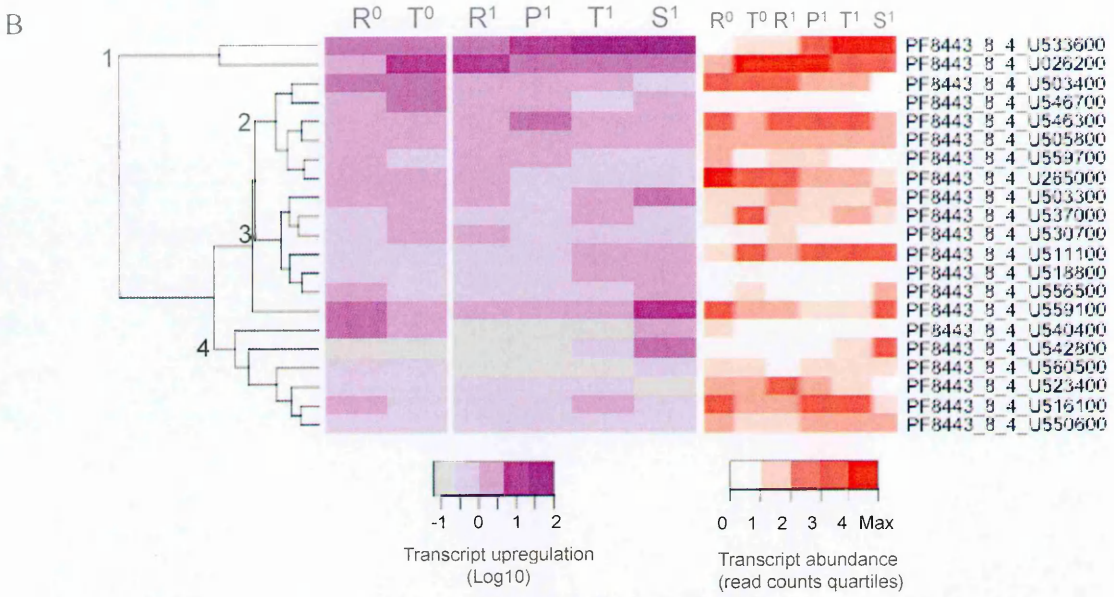


Figure 5.17. *Heatmaps showing rif (A) and stevor (B) transcript upregulation (purple) and transcript abundance (orange) in acute rings (R^0) and ex vivo trophozoites (T^0) and in cultured parasites at 4 IDC stages, ring (R^1), late ring to early trophozoites (P^1), trophozoite (T^1), and schizont (S^1). The left panel (purple) indicates log10 fold change transcript upregulation across the IDC. The right panel (orange) represents transcript abundance categorised into quartiles. The dendogram was generated by hierarchical clustering in R based on transcript upregulation pattern.*

5.6.2. *Rif* and *stevor* transcription in acute, culture- adapted, and culture-adapted and HBEC-selected 9106 lines.

For a second Kilifi isolate 9106, profiles were analysed for acute, culture-adapted and culture-adapted and HBEC-selected parasite lines. For this isolate RNA was not available for the ex vivo trophozoites, therefore only the acute ring transcription profile was analysed. There was insufficient *rif* sequence data for the HBEC-selected trophozoites due to problems during sequencing.

There was a higher diversity of *rifA* transcripts in acute parasites than in cultured or HBEC-selected parasites. Similar dominant *rifA* were observed in all lines in the ring stage, but there was a slight change in proportions. PF8443_8_7_U525700 was dominant in rings in all lines while PF8443_8_7_U439800 and PF8443_8_7_U484300 were predominantly transcribed in the HBEC-selected line. In the culture-adapted line there was a switch at the late ring-early trophozoite stage to a new dominant *rifA* PF8443_8_7_U503900 that stayed on in trophozoites and schizonts. No dominant rifA were unique to the HBEC-selected line, but two were dominantly transcribed in early stages in the HBEC-line with another two staying constant through the IDC (Figure 5.18).

In *rifB* the proportions of the dominant transcripts changed with the different selection pressures. PF8443_8_7_U443900, a pseudogene, was highly transcribed in rings in all the lines and in all the IDC stages of cultured line. In the HBEC-selected line this gene was expressed in much lower levels. PF8443_8_7_U499600 was most abundant in cultured rings and schizonts, and HBEC-selected schizonts. PF8443_8_7_U555300

was dominant in HBEC-selected rings and PF8443_8_7_U502300 was the dominant *rifB* in cultured and HBEC-selected late rings and early trophozoites (Figure 5.18).

Similar to what was observed for *rif* there was higher diversity of *stevor* transcripts in the acute rings than in the culture-adapted or HBEC-selected rings. Although the most abundant *stevor* transcripts were similar in the different lines, there were clear dominants uniquely associated with either the cultured or HBEC-selected lines. Three most abundant *stevor* in acute rings PF844_8_7_U536400, PF844_8_7_U503200 and PF844_8_7_U5557100 were also highly transcribed in cultured and selected parasites. PF844_8_7_U536400, the conserved stevor pseudogene (PFD0065w in 3D7), was predominantly transcribed across the IDC in the culture-adapted line. A different *stevor* PF844_8_7_U557100 not highly transcribed in culture was the most abundant transcript in HBEC-selected parasites. The culture-associated *stevor* PF844_8_7_U536400 was the second most abundant transcript in the HBEC-selected line.

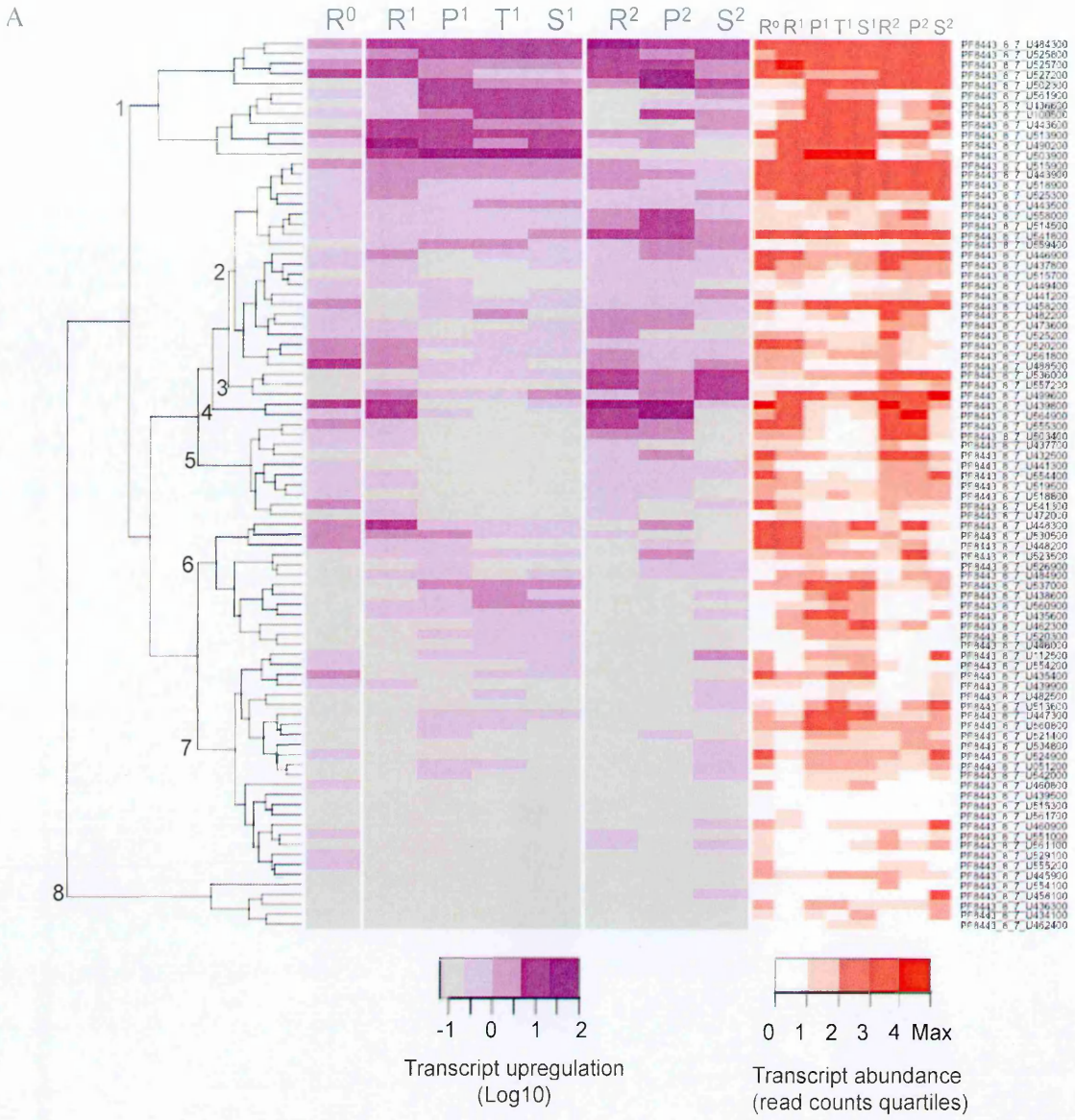
9106 IC & RD	Ring cDNA			LR / ET cDNA			Trophozoite cDNA			Schizont cDNA			Genomic DNA		
	rifA	*rifB*	*stev*	*rifA*	*rifB*	*stev*	*rifA*	*rifB*	*stev*	*rifA*	*rifB*	*stev*	*rifA*	*rifB*	*stev*
															
															
															

Label	*pir*	ID	A/B	BLAST top hit	Identities	2nd BLAST hit #	Identities
1	*rif*	PF8443_8_7_U525700	A	PFSAO75_128200	519/524 (99%)	PFIT_1400400	697/861 (80%)
2	*rif*	PF8443_8_7_U439800	A	PFIT_1401500	613/613 (100%)	PfHB3_BroadWash544900	973/976 (99%)
3	*rif*	PF8443_8_7_U484300	A	PFDG_02024.1|A	810/1010 (80%)	PF3D7_1200500 (PFL0025c)	501/582 (86%)
4	*rif*	PF8443_8_7_U555300	B	PF3D7_0324500 (PFC1100w)	832/955 (87%)	PFIT_0730900	825/949 (86%)
5	*rif* (c)	PF8443_8_7_U443900*	B*	PF8443_8_8_U484000*	660/660 (100%)	PF3D7_0533000 (PFE1635w)*	430/445 (96%)
6	*rif*	PF8443_8_7_U435400	B	PFIT_bin03600	872/879 (99%)	PfHB3_BroadWash529000	787/881 (89%)
7	*rif*	PF8443_8_7_U458200	B	PF8443_8_4_U547000	825/880 (93%)	PfHB3_BroadWash522400	805/880 (91%)
8	*rif*	PF8443_8_7_U503900	A	PF8443_8_4_U541400	752/885 (84%)	PF3D7_0833000 (MAL7P1.217)	727/867 (83%)
9	*rif*	PF8443_8_7_U446900	B	PF8443_8_8_U493000	921/936 (98%)	PFDG_00179	875/936 (93%)
10	*rif* (c)	PF8443_8_7_U499600	B	PF8443_8_8_U522100	945/1008 (93%)	PF3D7_0632300 (PFF1570w)	941/1012 (92%)
11	*rif* (c)	PF8443_8_7_U502300	B	PfHB3_BroadWash262300	953/957 (99%)	PF3D7_1000600 (PF10_0006)	936/957 (97%)
12	*rif*	PF8443_8_7_U564900	A	PF8443_8_3_U468800	1106/1107 (99%)	PF3D7_0937300 (PFI1805w)	1107/1116 (99%)
13	*stevor*	PF8443_8_7_U503200	na	PFIT_bin05800	777/908 (85%)	PfHB3_BroadWash526500	757/887 (85%)
14	*stevor* (c)	PF8443_8_7_U563400*	na	PF8443_8_1_U501600*	794/798 (99%)	PF3D7_0401500 (PFD0065w)*	454/457 (99%)
15	*stevor*	PF8443_8_7_U486000	na	PF3D7_1400700 (PF14_0007)	813/882 (92%)	PF8443_8_8_U514700	806/874 (92%)
16	*stevor*	PF8443_8_7_U557100	na	PFIT_0300100	810/908 (89%)	PF3D7_0115400 (PFA0750w)	782/909 (86%)
17	*stevor*	PF8443_8_7_U560500	na	PF8443_8_3_U442900	824/840 (98%)	PFIT_bin01300	816/840 (97%)

***Figure 5.18**. rif and stevor genomic repertoires and transcription profiling in 9106 at acute rings and ex vivo trophozoites (top panel), at 4 time points in culture-adapted parasites (middle panel) and HBEC-selected parasites (bottom panel). The pies represent proportions of reads that mapped to rifA, rifB or stevor. Coloured regions of the cDNA pies represent transcripts that accounted for 10% and above of the total expression in at least one time point, while in gDNA pies, coloured regions indicate genes whose read counts represent >10% of the total reads obtained from amplification of matched cDNA. The table (B) list the dominant transcripts (genes that took up at least 10% of total transcripts at any time point). !: piecharts from time points with insufficient sequence data.*

There were some clearly differentially upregulated *rif* in the different parasite lines, under different selection pressures (Figure 5.19. A). Cluster 1, which contained most of the dominant transcript was upregulated in cultured and HBEC selected lines. Cluster 3 was off in acute rings but upregulated in the cultured and selected lines. Clusters 4 and 5 were upregulated in acute rings, switched off in culture and turned back on after HBEC selection. Cluster 6 on the other hand was upregulated only in the cultured line. Cluster 2 was maintained on in the three lines while clusters 7 and 8 were mostly off.

A few *stevor* were differentially upregulated stevor transcripts (Figure 5.19. B). The dominantly transcribed *stevor* pseudogene in the cultured line PF8443_8_7_U563400 was upregulated in all lines (outlier on branch 1 Figure 5.19. B). PF8443_8_7_U560500 was upregulated in acute and HBEC selected rings but switched off in the culture-adapted parasites. PF8443_8_7_U557100 was exclusively upregulated in the middle stages in the HBEC-selected line. Cluster 5 came up only in the HBEC-selected parasites while cluster 7 was upregulated in acute rings and in the HBEC-selected line.


A
R0 R1 P1 T1 S1 R2 P2 S2
1
2
3
4
5
6
7
8
-1 0 1 2
Transcript upregulation
(Log10)
0 1 2 3 4 Max
Transcript abundance
(read counts quartiles)

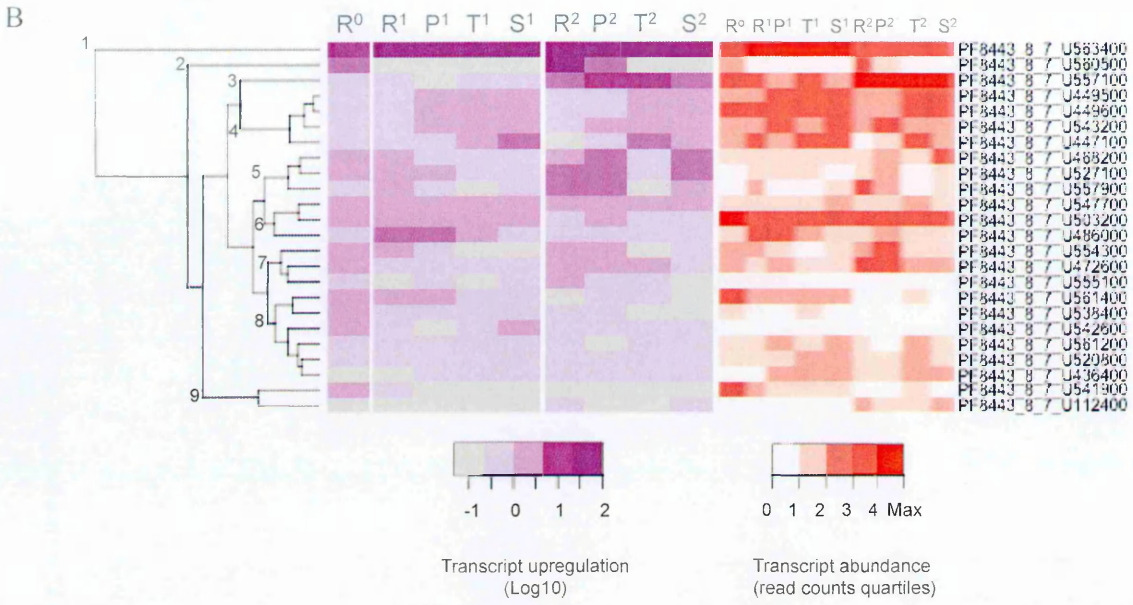


***Figure 5.19**. Heatmaps showing rif (A) and stevor (B) transcript upregulation (purple) and transcript abundance (orange) in a Kilifi isolate 9106, in acute rings (R^0), in the cultured line at 4 IDC stages, ring (R^1), late ring to early trophozoites (P^1), trophozoite (T^1), and schizont (S^1) and in the HBEC-selected line, ring (R^2), late ring to early trophozoites (P^2), trophozoite (T^2), and schizont (S^2). The left panel (purple) indicates log10 fold change transcript upregulation across the IDC. The right panel (orange) represents transcript abundance categorised into quartiles. The dendogram was generated by hierarchical clustering in R based on transcript upregulation patterns.*

5.6.3. *Rif* and *stevor* transcription in acute, culture- adapted, and culture-adapted and HBEC-selected 10668 lines.

The third Kilifi isolate 10668 was sampled from a severe malaria patient with impaired consciousness and respiratory distress. *rif* and *stevor* transcription profiles were analysed in the acute, cultured and HBEC-selected parasite lines. For this isolate there was insufficient *rif* sequence data for the ex vivo trophozoites due to problems during sequencing.

For this isolate there was high diversity of *rifA* transcripts in acute and HBEC-selected lines, with the cultured line showing a very homogenous transcription profile (Figure 5.20). The culture-adapted fraction had undergone three rounds of rosette selection using Gelatin (0.75%) floatation to increase its rosetting frequency. Rosette selection was stopped because there was no change in rosetting frequency, from 6.8% at acute to 4.5% after the third round of selection on the culture-adapted parasites.

The dominant *rifA* in acute rings PF8443_8_8_U533400 was transcribed in trophozoites and schizonts in the cultured line, and in all stages in HBEC-selected lines but at lower levels. In the culture-adapted parasites a single *rifA,* PF8443_8_8_U517800 dominated transcription in rings, early trophozoites and schizonts. In trophozoites a new dominant *rifA* PF8443_8_8_U293700 came up. This *rif* was transcribed in high levels in the HBEC-selected line, peaking in the trophozoite stage. Unlike for *rifA* there were multiple dominant *rifB* in rings all parasite lines. The two dominant *rifB* in acute rings PF8443_8_8_U 505900 and PF8443_8_8_U521100, were turned off in culture but were recovered after HBEC selection. There was a clear switch to a different dominant *rifB* between the cultured

and selected lines. While PF8443_8_8_U484000, a *rifB* pseudogene predominated transcription in cultured parasites PF8443_8_8_U521100 dominated transcription in HBEC-selected line, peaking in the trophozoite stage (Figure 5.20).

Multiple *stevor* transcripts were highly transcribed in acute rings, but a dominant transcript PF8443_8_8_U501200 was detected in cultured rings, and a different one PF8443_8_8_U542300 dominated in HBEC-selected rings. Very similar profiles were observed in the cultured and HBEC-selected lines in the middle stages of the IDC. One *stevor* PF8443_8_8_U561500 was predominant in the cultured line, while two other transcripts PF8443_8_8_U542300 and PF8443_8_8_U488800 were more abundant in the HBEC line (Figure 5.20).

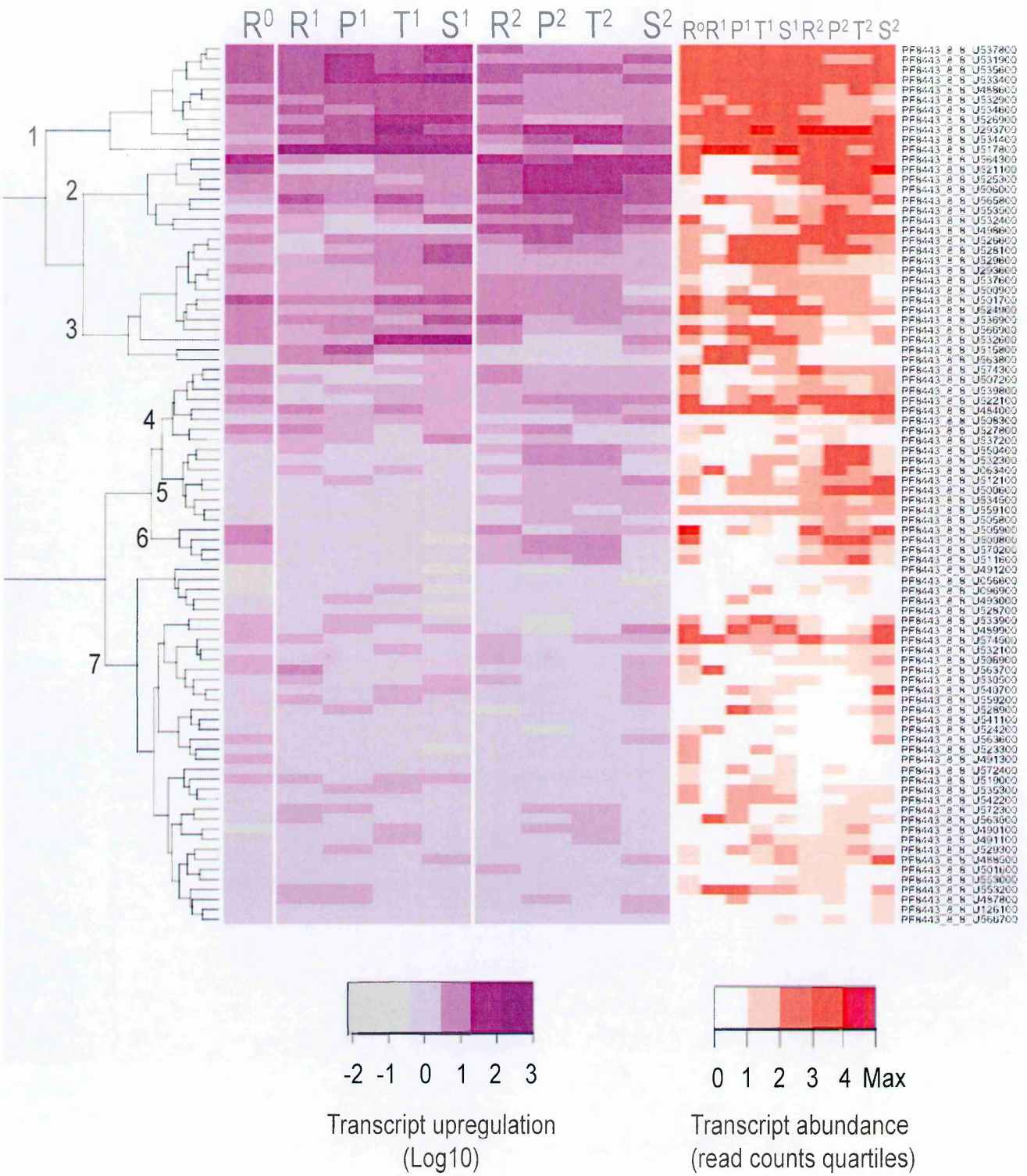
10668 IC & RD	Ring cDNA			LR / ET cDNA			Trophozoite cDNA			Schizont cDNA			Genomic DNA		
	rifA	*rifB*	*stev*	*rifA*	*rifB*	*stev*	*rifA*	*rifB*	*stev*	*rifA*	*rifB*	*stev*	*rifA*	*rifB*	*stev*
															
															
															

			A/B	BLAST top hit	Identities	2nd BLAST hit #	Identities
	rif	PF8443_8_8_U293700	A	PF8443_8_9_U142300	954/1000 (95%)	PF3D7_0808900 (PF08_0104)	952/999 (95%)
	rif	PF8443_8_8_U517800	A	PF8443_8_1_U550800	782/975 (80%)	PF3D7_0800500 (PF08_0138)	687/884 (77%)
	rif	PF8443_8_8_U533400	A	RIF_K401 (10595)	686/780 (87%)	PF3D7_1101300 (PF11_0021)	687/876 (78%)
	rif	PF8443_8_8_U505900	B	RIF_K433 (10595)	697/738 (94%)	PF3D7_0401000 (PFD0045c)	802/934 (85%)
	rif	PF8443_8_8_U521100	B	PFSAO75_493900	774/810 (95%)	DQ265331.1|Pf1235_47|B	737/791 (93%)
	rif	PF8443_8_8_U564300	B	PF3D7_0732700 (PF07_0136)	1064/1065 (99%)	PF8443_8_1_U552800	1064/1065 (99%)
	rif	PF8443_8_8_U574500	B	PFDG_00496	775/881 (87%)	PFIT_bin0480	767/876 (87%)
	rif (c)	PF8443_8_8_U484000*	B	PF8443_8_7_U443900	660/660 (100%)	PF8443_8_4_U512100	542/586 (92%)
	rif	PF8443_8_8_U534400	A	PF8443_8_6_U521600	878/981 (89%)	DQ265404.1|Pf1548_34|A	795/883 (90%)
	rif	PF8443_8_8_U532600	A	PF8443_8_10_U525500	816/975 (83%)	PF8443_8_7_U537000	740/890 (83%)
	rif	PF8443_8_8_U498600	A	PFDG_02024	951/1153 (82%)	PF8443_8_10_U566600	946/1154 (81%)
	rif	PF8443_8_8_U489900	B	PF3D7_0701200*	798/929 (85%)	PF8443_8_7_U445900	797/938 (84%)
	stevor	PF8443_8_8_U535700	na	PF3D7_1254300 (PFL2620w)	832/910 (91%)	PFSAO75_154400	825/909 (90%)
	stevor	PF8443_8_8_U488800	na	PFDG_00592.1	823/898 (91%)	PF3D7_147990 (PF14_0771)	814/916 (88%)
	stevor	PF8443_8_8_U542300	na	PFDG_00092.1	645/660 (97%	PF8443_8_8_U097100	626/662 (94%)
	stevor	PF8443_8_8_U553300	na	PF8443_8_4_U503300	825/873 (94%)	PF3D7_0400800 (PFD0035c)	773/864 (89%)
	stevor	PF8443_8_8_U501200	na	PF8443_8_10_U525300	846/846 (100%)	PFDG_00184	580/612 (94%)
	stevor	PF8443_8_8_U563300	na	PF8443_8_4_U265400	567/567 (100%)	PFIT_bin00600	730/796 (91%)
	stevor	PF8443_8_8_U561500	na	PF8443_8_10_U542900	811/922 (87%)	PF8443_8_4_U511100	579/612 (94%)

***Figure 5.20**. rif and stevor genomic repertoires and transcription profiling in 10668 at acute rings and ex vivo trophozoites (top panel), at 4 time points in culture-adapted parasites (middle panel) and HBEC-selected parasites (bottom panel). The pies represent proportions of reads that mapped to rifA, rifB or stevor. Coloured regions of the cDNA pies represent transcripts that accounted for 10% and above of the total expression in at least one time point, while in gDNA pies, coloured regions indicate genes whose read counts represent >10% of the total reads obtained from amplification of matched cDNA. The table (B) list the dominant transcripts (genes that took up at least 10% of total transcripts at any time point). !: piecharts from time points with insufficient sequence data.*

Certain *rifs* were preferentially upregulated under the different selection pressures. *The rifs* in cluster 1 were upregulated in acute and cultured lines, with marginal upregulation in the HBEC-selected line. The most abundant *rifA* were clustered in this group, including the dominant *rifA,* PF8443_8_8_U517800 in the early stages in the culture-adapted line. Cluster 2, which contained two of the most abundant *rifB* was preferentially upregulated in the middle stages of the IDC in HBEC-selected parasites. Cluster 4 was mostly up in acute while clusters 5 came up after HBEC selection. Cluster 6 was on in acute and HBEC-selected parasites (Figure 5.21. A).

For stevor few transcripts were significantly upregulated. *Stevors* in cluster 1 were upregulated in acute, cultured and HBEC-selected lines. Cluster 2 was off in acute and culture-adapted lines but upregulated after HBEC-selection. Cluster 3 was most upregulated in acute, stayed on in cultured parasites and switched off in the middle stages of the HBEC line. Cluster 6 was predominantly upregulated in acute parasites Figure 5.21. B).


R0 R1 P1 T1 S1 R2 P2 T2 S2
1
2
3
4
5
6
7
-2 -1 0 1 2 3
Transcript upregulation
(Log10)
0 1 2 3 4 Max
Transcript abundance
(read counts quartiles)

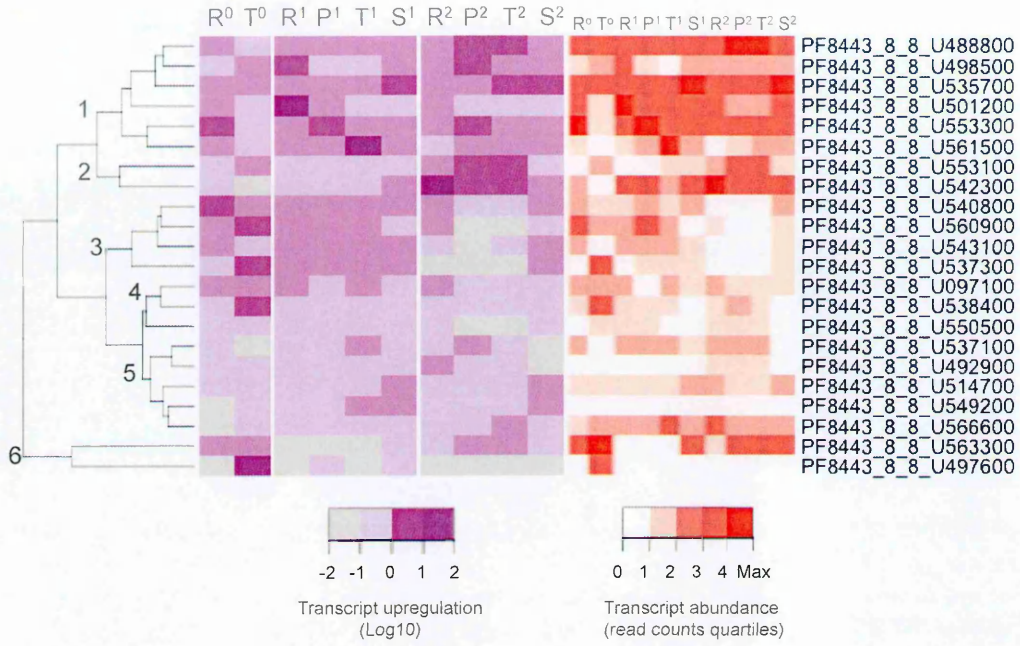


Figure 5.21. *Heatmaps showing rif (A) and stevor (B) transcript upregulation (purple) and transcript abundance (orange) in a Kilifi isolate 10668, in acute rings (R^0) and ex vivo trophozoites (T^0), in the cultured line at 4 IDC stages, ring (R^1), late ring to early trophozoites (P^1), trophozoite (T^1), and schizont (S^1) and in the HBEC-selected line, ring (R^2), late ring to early trophozoites (P^2), trophozoite (T^2), and schizont (S^2). The left panel (purple) indicates log10 fold change transcript upregulation across the IDC. The right panel (orange) represents transcript abundance categorized into quartiles. The dendogram was generated by hierarchical clustering in R based on transcript upregulation patterns.*

5.6.4. *Rif and stevor transcription in a rosetting culture-adapted wild isolate SA075.*

Analysis of *rif* and *stevor* transcription in a rosetting culture-adapted wild isolate SA075 was carried out to determine whether any *rif/stevor* were associated with rosetting phenotype. The highly rosetting isolate was sampled from a patient with severe malaria anaemia residing in Kisumu, an area of high malaria transmission. The acute sample was not available for this study therefore transcription profiling was done for the rosetting clone (R+) and the non-rosetting uncloned fraction (R-) at 4 IDC stages.

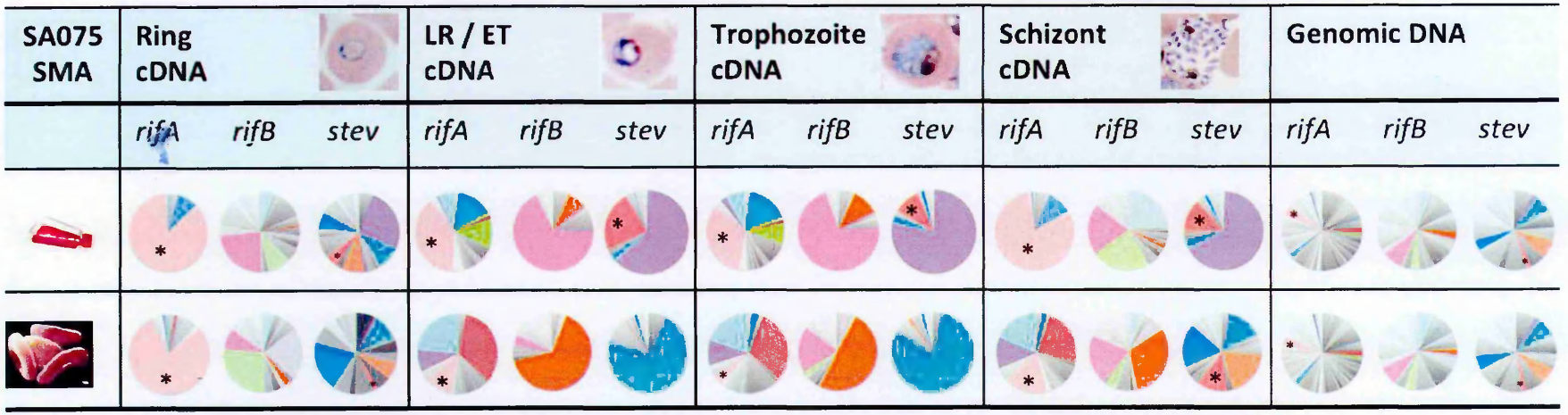
A *rifA* pseudogene PFSA075_PN48 was dominantly transcribed in rings in both the rosetting and non-rosetting lines, and was sustained through the IDC in the non-rosetting fraction, with a second peak in the schizont stage. In the middle stages PFSA075_071800 came up in the R- while PFSA075_128200 and PFSA075_PN57 dominated transcription in the R+ trophozoites and schizonts. The pseudogene PFSA075_PN48 was also highly transcribed in the R+ schizonts (Figure 5.22).

There was an observable difference in the *rifB* transcribed by the different parasite lines especially in the middle of the IDC (Figure 5.22). In rings the dominant transcript was PFSA075_PN19 in R- and PFSA075_494000 in R+. PFSA075_PN19 stayed dominant in R- late rings to trophozoites decreasing in abundance in schizonts. PFSA075_220500 was the dominant R+ late rings/early trophozoites, trophozoites and schizonts (Figure 5.22).

Stevor transcription profile was similar to *rifB* profile with a marked switch in the middle stage to a different dominant *stevor* after rosette selection. PFSA075_128400

was the dominant transcript in R+ late rings to trophozoites while PFSA075_493100 was dominant in R- (Figure 5.22).

Similar results were previously obtained in the capillary data showing differential expression of *rif* and *stevor* in rosetting and non-rosetting parasites especially at trophozoite stage (Chapter 4). Comparing the two data sets revealed a difference in transcripts abundance of certain genes, due to the difference in coverage by the two methods but overall the dominant transcripts observed in the different parasite lines were similar in the two data sets.



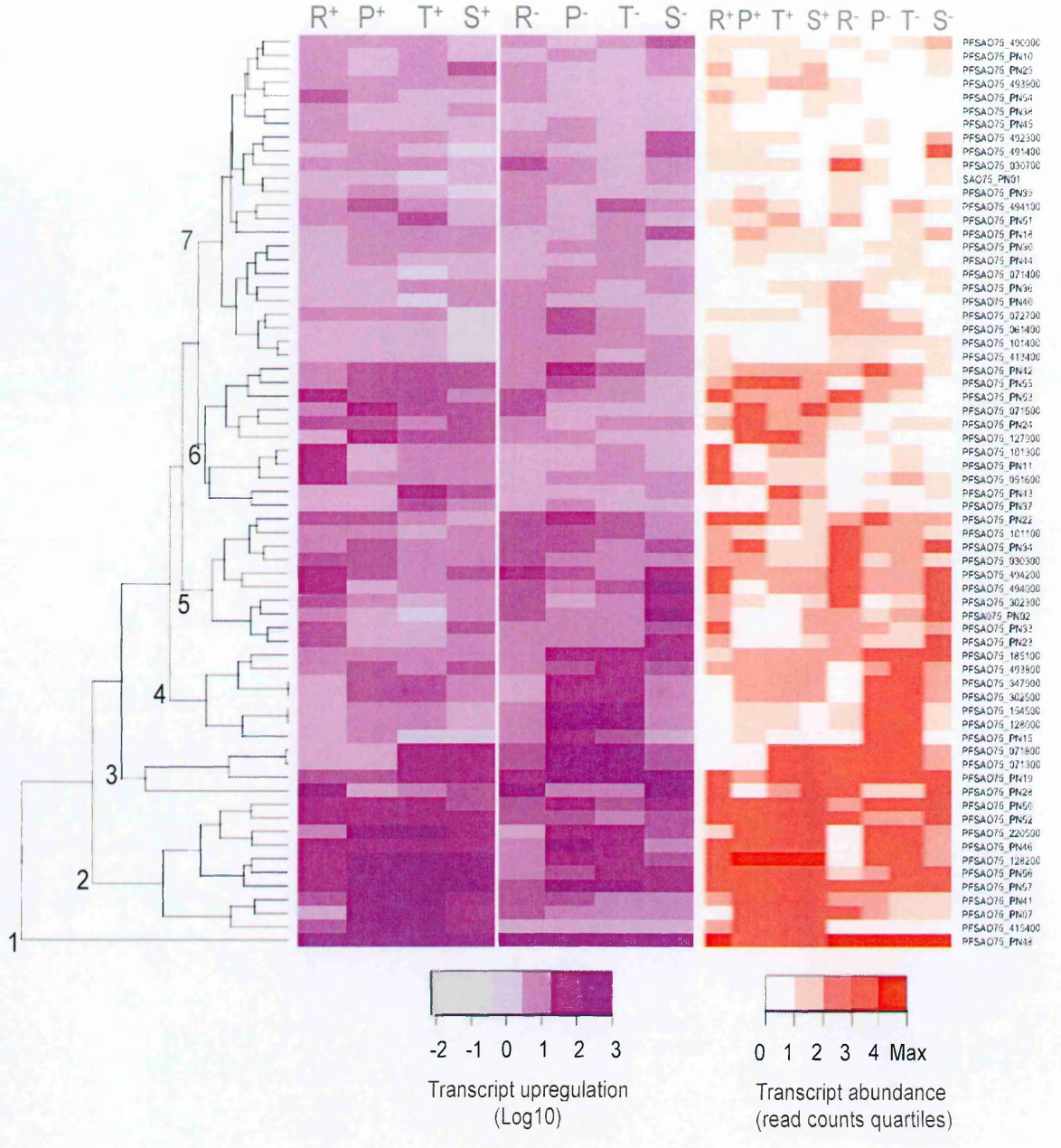
colour	rif/stevor	ID	A/B	Top BLAST hit	Identities	2nd BLAST hit #	Identities
	rif	PFSAO75_PN48*	A*	PF3D7_1372700 (MAL13P1.500)	912/997 (91%)	PF8443_8_1_U517500	913/997 (91%)
	rif	PFSAO75_PN57	A	PF3D7_0833400 (MAL7P1.213)	793/870 (91%)	PFIT_0537500	795/893 (89%)
	rif	PFSAO75_128200	A	PF3D7_0700200 (MAL8P1.219)	799/892 (89%)	PF8443_8_10_U507300	778/901 (86%)
	rif	PFSAO75_PN56	A	PF8443_8_1_U553600	481/527 (91%)	PF3D7_0900200 (PFI0010c)	456/506 (90%)
	rif	PFSAO75_071800	A	PF8443_8_7_U100500	930/993 (93%)	PF3D7_0617700 (PFF0855c)	903/1005 (89%)
	rif	PFSAO75_PN28	A	PF3D7_0400700 (PFD0030c)	837/1024 (81%)	PF3D7_0800500 (PF08_0138)	825/1033 (79%)
	rif	PFSAO75_154500	A	RIF_K123 (10735)	757/860 (88%)	PfHB3_BroadWash527800	645/740 (87%)
	rif	PFSAO75_PN19	B	PF3D7_0222500	764/875 (87%)	PF8443_8_3_U569900	758/876 (86%)
	rif	PFSAO75_101300	B	PF3D7_0631800	849/1031 (82%)	PF8443_8_1_U220500	348/393 (88%)
	rif (c)	PFSAO75_220500	B	PF8443_8_1_U554200	1074/1086 (98%)	PF3D7_1000600 (PF10_0006)	1025/1087 (94%)
	rif	PFSAO75_PN02	B	DQ265347.1	782/783 (99%)	739_2_7a (10739)	714/716 (99%)
	rif	PFSAO75_494000	B	DQ265344.1	802/825 (97%)	PF8443_8_8_U540700	778/868 (89%)
	stevor	PFSAO75_154400	na	PF8443_8_8_U540800	839/909 (92%)	PF3D7_1254300 (PFL2620w)	833/909 (91%)
	stevor	PFSAO75_493100	na	PF8443_8_7_U43640	850/914 (92%)	PF3D7_0631900 (PFF1550w)	849/914 (92%)
	stevor	PFSAO75_128400	na	PF8443_8_1_U515000	847/906 (93%)	PF3D7_0222800 (PFB1020w)	831/915 (90%)
	stevor	PFSAO75_051700*	na	PF3D7_0401500 (PFD0065w)*	927/933 (99%)	PFIT_0400200*	926/933 (99%)
	stevor	PFSAO75_001200	na	PF3D7_1149900 (PF11_0516)	853/927 (92%)	PfHB3_BroadWash538600	804/928 (86%)

Figure 5.22. *rif and stevor genomic repertoires and transcription profiling in SA075 rosetting clone (top panel) and non-rosetting fraction (bottom panel) at 4 time point across the IDC. The pies represent proportions of reads that mapped to rifA, rifB or stevor Coloured regions of the cDNA pies represent transcripts that accounted for 10%and above of the total expression in at least one time point, while in gDNA pies, coloured regions indicate genes whose read counts represent >10% of the total reads obtained from amplification of matched gDNA. The table (B) list the dominant transcripts (gene that took up at least 10% of total transcripts at any time point). *: pseudogene.*

Differential transcript upregulation was observed for SA075 *rif*. PFSA075_PN48 falls out of the clusters because it is the only *rif* that was strongly upregulated (over 100 fold upregulation) in both the R+ and R- parasites. Clusters 2 and 3 were mostly upregulated in the R+ parasites with a slight upregulation in R- late rings/ early trophozoites and trophozoites. Cluster 4 was upregulated in the middle stages of the R- parasites. Cluster 5 was upregulated in R+ rings and R- schizonts. Cluster 6 was slightly upregulated in R+ more than in R-. Overall abundance and upregulation of *rif* transcripts were correlated (Figure 5.23. A).

stevor cluster 1 was uniquely upregulated in the non-rosetting fraction while cluster 2 was predominantly upregulated in the rosetting clone. Cluster 6 was only slightly upregulated in the rosetting clone (Figure 5.23. B).

A


R+ P+ T+ S+ R- P- T- S-
R+P+ T+ S+ R- P- T- S-
1
2
3
4
5
6
7
-2 -1 0 1 2 3
Transcript upregulation
(Log10)
0 1 2 3 4 Max
Transcript abundance
(read counts quartiles)

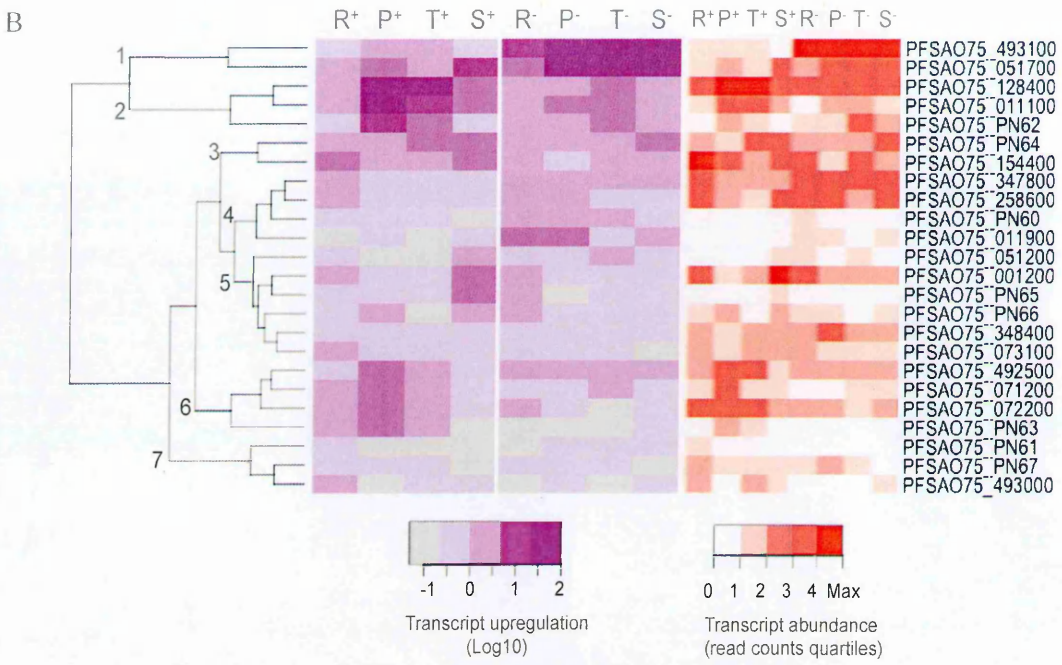


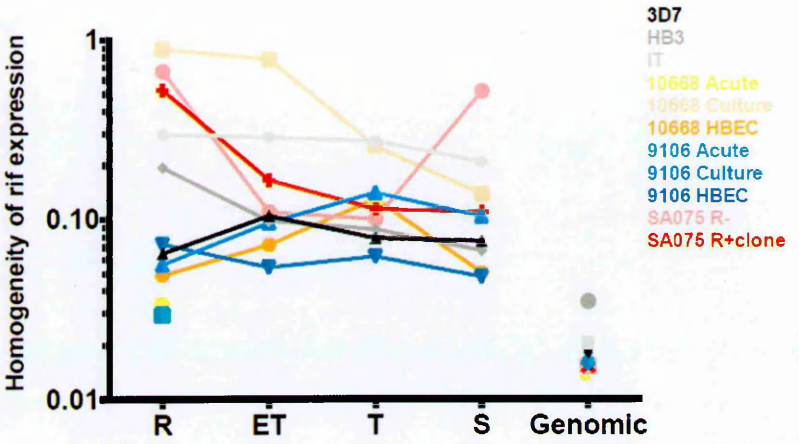
Figure 5.23. *Heatmaps showing rif (A) and stevor (B) transcript upregulation (purple) and transcript abundance (orange) in a rosetting Kisumu isolate SA075, in the rosetting clone (+) and the non-rosetting fraction (-) in 4 IDC stages, ring (R^+)/(R^-), late ring to early trophozoites (P^+)/ (P^-), trophozoite (T^+)/ (T^-), and schizont (S^+)/ (S^-). The left panel (purple) indicates log10 fold change transcript upregulation across the IDC. The right panel (orange) represents transcript abundance categorised into quartiles. The dendogram was generated by hierarchical clustering in R based on transcript upregulation patterns.*

5.7. Analysis of *pir* transcript diversity across the IDC in laboratory and field isolates.

We used the Simpson's diversity index to analyse homogeneity of *rif* expressed across the IDC in different parasite lines. There was no clear pattern in transcript diversity across the isolates (Figure 5.24). As observed in the pie charts acute parasites showed the least homogeneity in *rif* expression. There was no consistent pattern in transcript diversity by stage. Some parasites, culture-adapted 10668, HBEC-selected 9106, SA075 R+ clone, IT and HB3, showed highest homogeneity in rings while others, 3D7, HBEC-selected 10668 and culture-adapted 9106 showed highest homogeneity in the mid IDC stages. SA075 R- fraction had a divergent pattern with the least transcript diversity observed in rings and schizonts (Figure 5.24).

For *stevor* highest diversity was observed in rings with the middle stages having the least diversity. In schizonts there was a general decrease in homogeneity. Overall acute parasites showed higher diversity of stevor transcripts than the corresponding culture-adapted and cultured and HBEC-selected lines (Figure 5.24).

A



B

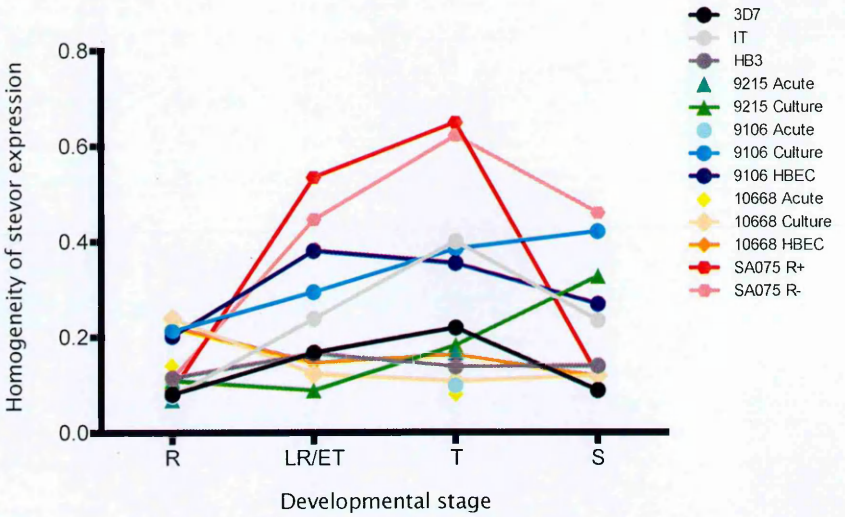


Figure 5.24. *Simpson's diversity index for rif (A) and stevor (B) transcripts across the IDC in laboratory and wild isolates.*

5.8. Discussion

Expression patterns of *rif* and *stevor* genes have been analysed in laboratory isolates and in a few wild isolates and the evidence showed these genes to be expressed in multiple stages of the parasite life cycle, importantly in multiple stages of the IDC. These genes are not under mutually exclusive expression with multiple transcripts detected in a population and for *stevor* multiple transcripts have been detected in a single IE (Kaviratne et al., 2002). One study that carried out in depth expression profiles in 5 patient isolates revealed that there were 2 stages of peak VSA expression at rings and trophozoites, with the same *pir* variants in rings and trophozoites (Bachmann et al., 2012).

For other members of the *pir* multigene family, *cir*, *bir* and *yir*, *vir*, expression studies have revealed that multiple variants are expressed in parasite populations, and in individual IE (Fernandez-Becerra et al., 2005, Cunningham et al., 2009, Cunningham et al., 2010). A recent study on phylogeny and expression of *pir* from rodent malaria parasites showed that distinct groups of *pirs* were expressed in asexual and sexual stages of *P. berghei*, suggesting functional divergence (Otto et al., 2014a). In *P. chabaudi* A and B *cir* subgroups were expressed at similar levels in parasite populations, although the dominant transcripts belonged to the A subgroup (Lawton et al., 2012). In another study, mosquito transmitted parasites showed a more even transcription of many *cir* genes while continuously blood passaged parasites selected for expression of a dominant *cir* and this resulted in increased virulence (Spence et al., 2013). This showed a direct link between pir gene expression and disease virulence.

To understand *P. falciparum pir* gene expression, temporal transcription profiles were explored in laboratory and wild *P. falciparum* isolates by deep amplicon sequencing. Dynamic expression patterns in the different isolates were observed. Clear temporal expression patterns were observed, with certain genes dominating expression at certain stages of the IDC. Different dominant transcripts were observed at the early (ring), mid (trophozoite) and late (schizont) IDC stages. The clear temporal expression pattern suggests that some *rif* and *stevor* function at specific stages of the IDC. For instance, the genes transcribed in the early and mid trophozoites might be expressed on the surface of the iRBC, while the transcripts that come up in late trophozoites and schizonts might be expressed on the surface of merozoites and play a role in immune evasion of the merozoites or during invasion. A previous study identified two *rif* transcripts, PF3D7_0900500 (PFI0025c) and PF3D7_1300600 (PF13_0006), that were expressed in the early (early ring) and late (schizont) stages of the parasite's IDC (Wang et al., 2009). In a later study PF3D7_1300600 was shown to be expressed in late stage schizonts and the protein detected in merozoites and gametes (Mwakalinga et al., 2013). The early transcripts could be carried over mRNA from late stages of the parasite IDC. The question as to whether or not these transcripts are translated into proteins is remains to be explored. There is evidence that mRNA half-life increases as parasites mature (Shock et al., 2007), mRNA from early stages is degraded much faster than in later stages. Whether this has an effect on protein expression levels is still unclear.

In the wild isolates the temporal expression pattern was observed but in addition, differential expression of *pir* was observed under the different selection pressures. For *rifA* there was more even expression in ex vivo parasites, while culture-adapted and HBEC-selected or rosette-selected parasites had more focused expression to a few

dominant variants. For *rifB* and *stevor* the differential expression profiles was more evident, with distinct dominant variants in ex vivo parasites, and in culture-adapted and parasites selected for binding phenotypes (HBEC-binding and rosetting). A noteworthy observation was the differential transcription of *rif* and *stevor* in the rosetting trophozoites. There is evidence of upregulation of certain *rif* genes after selection for rosetting or binding to HBEC, but this is as a result of linked expression between UpsA *var* and their neighbouring *UpsArif* that share promoter regions (Claessens et al., 2011, Claessens et al., 2012). In our case the dominant *rif* genes in selected lines did not belong to the *UpsArif* group, by BLAST analysis, and did not lie in close proximity to the upregulated *var,* although this will need to be verified after correction and manual curation of genome assemblies of the wild isolates is done. This suggest that *rif* and *stevor* upregulation was independent of *var* expression in the patient isolates. A recent study has shown that *rifA* (Goel et al., 2015) and *stevor* mediate rosetting (Niang et al., 2014) and it is plausible that *pir* genes are involved in cytoadhesion. The dominant *stevor* in 3D7 PF10_0395 was previously shown to bind to human RBC and mediate rosetting in A4 in the aforementioned study (Niang et al., 2014). Our 3D7 parasite line did not form rosettes to any notable degree, implying that other factors come into play to result in a specific adhesion property, including post transcription and translation regulations, and protein translocation to iRBC surface.

The number of transcripts detected even in laboratory-adapted lines was notably high, suggesting an important role for post-transcriptional modifications in regulation of *pir* expression. Posttranscriptional regulation for mutually exclusive expression was recently

demonstrated for upsA *var* genes, which involved degradation of nascent RNA by the newly discovered exoribonuclease PfRNaseII (Zhang et al., 2014). The detection of pseudogenic transcripts, some dominantly expressed in some isolates raise the question of the role of pseudogenes and whether pseudogenes may play a role in gene expression regulation.

There was a bias towards *rifA* transcription in laboratory isolates and in 3 out of 4 culture-adapted wild isolates. In the wild isolates there was an increase in *rifB* transcripts in the ex vivo and HBEC-selected parasites lines. This finding was unexpected because according to the current dogma *rifB* do not play a role in host-parasite interactions, and are not trafficked to the surface of iRBC. Therefore the expectation was that *rifB* would be downregulated in ex vivo parasites. Previous studies suggested that *rifA* and *rifB* had diverged functionally, with *rifA* playing a role in host-parasite interactions on the surface of the iRBC and *rifB* staying inside the PVM (Petter et al., 2007). In a case study of one splenectomized patient it was shown that all VSA except *rifB* and *pfmc2tm* were switched off and mature parasites were found in peripheral circulation (reflecting lack of cytoadhesion). These parasites did not bind the common endothelial receptors suggesting that the *rifB* and pfmc2tm do not play a role in adhesion to endothelial cells (Bachmann et al., 2009). Our data indicates that *rifB* were upregulated in ex vivo parasites and selected lines, but turned off in culture, suggesting a function within the host environment. In addition there was restricted diversity of *rifB* transcripts, with a single or few dominantly expressed. Further studies on localisation and binding might elucidate the

role of these highly expressed *rifB,* after validating their upregulation either by whole transcriptome sequencing or quantitative realtime PCR.

In line with previous studies, upregulation of UpsA *var* genes in 3D7 was linked with the upregulation of the neighbouring *rif* gene in rings. This could be as a result of switching on of the strong *var* promoter, because the transcription pattern of the *rifA1* was similar to that of *var*. It has been shown that *UpsA* associated *rif* are upregulated in parasites that express their associated *var* (Wang 2009, Claessens 2012). Three members of this conserved group of *rifs* in 3D7 PF13_0004, PF11_0009 and PF11_0520 were the most prominent transcripts in 3D7 rings, but not in the mid and late stages. From the transcript upregulation data, the UpsA *rif,* PFD1230c, was the most upregulated *rif* in 3D7 rings. The neighbouring *var* PFD1235c that is in head-to-head orientation with this *rif* was the second most dominantly expressed *var* gene by tag analysis (Andisi et al., personal communication). In fact three of the four UpsA *rif* recovered in our analysis cluster together in cluster 4 by transcript upregulation pattern (Figure 5.10) suggesting they have similar transcription patterns, with highest upregulation in the early stages, lowest in mid and late trophozoites and slight increase in schizonts. The observed increase in transcript upregulation in schizont stage is most likely due to contamination of the schizonts with rings, as the cultures lost synchrony towards the end of the IDC (see Appendix table 9.5, parasite staging).

A highly transcribed IT pseudogene lies in a conserved region between a conserved *rif* and the conserved *var1csa. In* 3D7 this region on chromosome 5 is also conserved with a

conserved *rif* PFE1630w and the *var1csa* PFE1640w. The fact that the *rif* pseudogene is transcribed in the early stages of the IDC, similar to *var* transcription timing raises the question whether there is a link in transcription for genes in close proximity even though they are not in head-to-head orientation. Epigenetic regulation of gene expression may play a role. Studies on transcription regulation have shown that nuclear positioning is a factor for *var* gene transcription regulation and this might be true for *pir* genes.

It is important to note that isolates that were selected for rosetting phenotype by gelatin floatation had a very restricted *rif* transcript diversity, with a few transcripts dominating in certain stages of the IDC. In IT, three *rifA* transcripts dominated most of the IDC, a pseudogene in the early stages, and two intact *rifA* in the mid and late stages. A single *rifB* dominated from late rings to schizonts. In 10668 cultured line that underwent rosette selection (but never displayed any significant rosetting capacity), a single *rifA* dominated transcription in the first half of the IDC. In SA075, for both the rosetting and non-rosetting parasites, one *rifA* dominated the early stages. *rifB* transcription was restricted to a single dominant transcript in the mid and late stages. Similarly *stevor* transcription was also restricted to some extent, after selection using gelatin. For the few isolates in our study that underwent selection, it appears as though contact with gelatin modified *rif* and *stevor* transcription. This is from the observation that all parasite isolates that underwent rosette selection using gelatin had a switch in preference to *rifA* expression (Figure 5.8. C, F and G). Our rosette enrichment protocol utilizes gelatin floatation to separate out the high density rosettes already formed in culture. This switch by selected parasites to predominantly expressing *rifA* could be as a result of parasites coming into contact with gelatin itself, or as a result of selection of a subpopulation of the

parasites that formed rosettes. This effect will need to be validated further with other isolates.

A major limitation of this study was our inability to capture the complete *rif* and *stevor* repertoire using the degenerate primers. The primers, which were designed on 3D7 genome, performed much better in Kenyan isolates and IT that they did in HB3. Still a minor proportion of the *rif* and *stevor* repertoire was not captured. Therefore there still remains a gap in describing transcription of *rif* and *stevor* in wild isolates that can only be filled using transcriptomics. This study has provided a first exploration of *rif* and *stevor* transcription in patient isolates under different selection pressures.

CHAPTER 6

6. De novo assembly of *rif* and *stevor* amplicon sequencing reads: An alternative analysis of 454 sequence reads in the absence of whole genome data.

6.1. Introduction

This chapter follows on from Chapter 5 on analysis of *pir* (*rif* and *stevor)* amplicon sequence data generated using the 454 sequencing platform. The focus of this chapter is the *de novo* analysis of 454 sequence data by contig assembly. The data presented was generated from *pir* amplicons from *P. falciparum* field isolates that lacked whole genome sequences. Therefore analysis of reads by directly mapping them onto a homologous genome was not possible. Because of suboptimal 454 sequencing runs (Figure 5.3) many reads obtained were short and did not span full-length *rif* or *stevor* amplicons (Figure 5.2). De novo assembly of reads into contigs that would potentially cover the full-length *pir* amplicon was done using the Newbler assembly package. Newbler assembler merges short reads to form a long contig by identifying the best sequence overlaps between the reads.

The *pir* genes are highly variable, especially the *rif* gene family, which has few conserved orthologs across different parasite repertoires. The *stevor* genes are relatively conserved especially the major group *stevorA* (as discussed in Chapter 3). For this reason analysis of next generation sequence data by mapping onto non-homologous reference genomes could be inaccurate, as only a small proportion of *pirs* would map. Whole genome sequence data from wild parasite isolates is now available from a large global collection

of isolates, but the VSA data from these is still sparse because of the difficulty in assembly of short NGS correctly into full length VSA. For this reason laboratory lines such as 3D7 are still used as reference. Therefore analysis of *pir* expression from patient isolates is inherently difficult. De novo assemblies of 454 read data provided a robust alternative approach to analyse *rif* and *stevor* expression in wild isolates.

6.2. Objectives

i.) Optimize protocol for *de novo* contig assembly of *pir* reads from 454-amplicon sequencing.

ii.) Account for sequence errors and length differences in the *de novo* assembled contigs using a combination of CD-HIT clustering and BLAST.

iii.) Analysis of *pir* expression profiles through a single *ex vivo* blood stage cycle in clinical isolates.

6.3. Methods

The methods used in this chapter are described in Chapter 2 and in Chapter 5 Methods. The data analysed in this chapter was obtained by 454 amplicon sequencing of *pir* genes from 3D7 and patient isolate IDC time series samples. The parasite isolates used are described in Chapter 5, Figure 5.1. and Table 5.1. In brief, 3D7 time course RNA was extracted from synchronous cultures at 6 time points across a single 48-hour blood stage cycle, and gDNA from asynchronous cultures. 8383, a culture-adapted Kilifi isolate with poor whole genome sequence data had been obtained from an ongoing study on *var* expression led by Dr. Peter Bull. RNA had been extracted from isolate 8383 rings sampled at the acute stage of infection, and trophozoites, sampled 24 hours after growth in culture (ex vivo). In addition, RNA was extracted from four asexual stages of the culture-adapted 8383 isolate, rings (R), late rings to early trophozoites (LRET), mid to late trophozoites (T) and schizonts (S). For the Kilifi *ex vivo* isolates the time course experiment was carried out by culturing the isolates for a single IDC. RNA was extracted at 7 time points every 10 hours across the IDC (Appendix table 9.5). The parasites and time points for which *pir* amplicon sequencing was done are summarized in Table 6.1. Preparation of cDNA was carried out as described in Chapter 2, sections 2.6.1 and 2.6.2. Purified amplicons were sequenced using the 454 sequencing platform at WTSI as described in Chapter 2, section 2.8. For many of the *ex vivo* patient isolates only limiting amounts of RNA and DNA were available for *pir* amplification and sequencing. For this reason many samples had to be artificially loaded with carrier DNA of bovine origin (from the 454 sequencing team at the Sanger Institute) to make up to required starting amplicon quantities of 100ng total DNA per sample.

6.3.1. Newbler assembly of 454 reads

Assembly of 454 reads into contigs was carried out using Newbler assembler version 2.5.3 in a script (done by Dr. Etienne de Villiers, KWTRP head of bioinformatics, See Appendix, script 9.1). The assembly was run using default newbler parameters (appendix 9.1). In summary, the following basic parameters were used, minimum read length of 20 nucleotides for the input, a minimum overlap length of 40 nucleotides with 90% identity within the overlap region. The minimum input length of 20 nucleotides was not altered in the newbler script because the restriction on the length of overlap, 40 nucleotides, as well as the minimum output length of 100 nucleotides would ensure that we obtained contigs longer than 100 nucleotides for downstream analysis.

6.3.2. Cleaning up de novo assembled contigs.

The *de novo* assembled contigs contained many non-target sequences. These non-*pir* sequences were made up of contaminants from non-specific PCR amplifications (arising because of the highly degenerate primers used to capture the diverse *pir* gene families), bovine carrier DNA (added to amplicons that did not meet the required concentration for sequencing), human and bacterial DNA.

For the *rif* contigs, these non-target sequences were identified by a combination of multiple sequence alignments, clustering using networks and the CD-HIT algorithm (Li and Godzik, 2006), and a BLAST analysis of all contigs against NCBI nr nucleotide database and our *rif* database containing *rif* sequences from sequenced genomes and partial repertoires from published and unpublished data, and from our capillary

sequencing data (described in Chapter 4). For *stevor* an initial BLAST search against NCBI nucleotide nr database enabled identification of non-stevor contigs.

6.3.3. Clustering *rif* and *stevor* contigs

An online web server CD-HIT suite (http://weizhong-lab.ucsd.edu/cdhit_suite/cgi-bin/index.cgi) (Huang et al., 2010) was used for clustering of contigs to obtain unique *rif* and *stevor* variants. The CD-HIT clustering algorithm used is based on greedy incremental clustering. The contigs were first sorted by length with the longest sequence designated as the cluster representative. All sequences were then compared to the first cluster representative. If the similarity was above that of the set threshold then the sequence would be grouped into that cluster. If the similarity fell below the set threshold the sequence would become the representative of the next cluster. Short word filtering was used to determine similarity between sequences. If sequence similarity could not be confirmed by short word filtering then an actual alignment was done (Huang et al., 2010, Li and Godzik, 2006). Further clustering was done to collapse sequences into only unique contigs. The cluster representatives from CD-HIT clustering were searched against *rif* and *stevor* databases by MEGABLAST in Geneious v 6.1.5. Sequences with a BLAST hit above 96% pairwise identity were grouped into one cluster.

6.4. Results

6.4.1. 454 amplicon sequence reads assembly.

All the reads from the successful 454 sequencing runs (563,980 reads from *rif* amplicon sequencing and 793,118 reads from *stevor* amplicon sequencing) were assembled into contigs using Newbler software carried out in house. The parameters with which the assembly was run are described in the methods section above. Assembly was carried out for each isolate per sample (time point). Therefore for each time point a set of contigs was obtained. The contigs were named to reflect the isolate name, gene (*rif* or *stevor*), isolate sample type, whether acute (AC), cultured (CU), or selected for binding to human brain endothelial cell lines (HB), time point of RNA sampling, stage and sequence run. For example the first *rif* contig from 3D7 early ring sample would be named PF3D7_RIF_CU_00_ER_3_1. A total of 8246 and 2601 contigs from *rif* and *stevor* amplicon sequencing respectively were assembled (Table 6.1 and Figure 6.2 and 6.3).

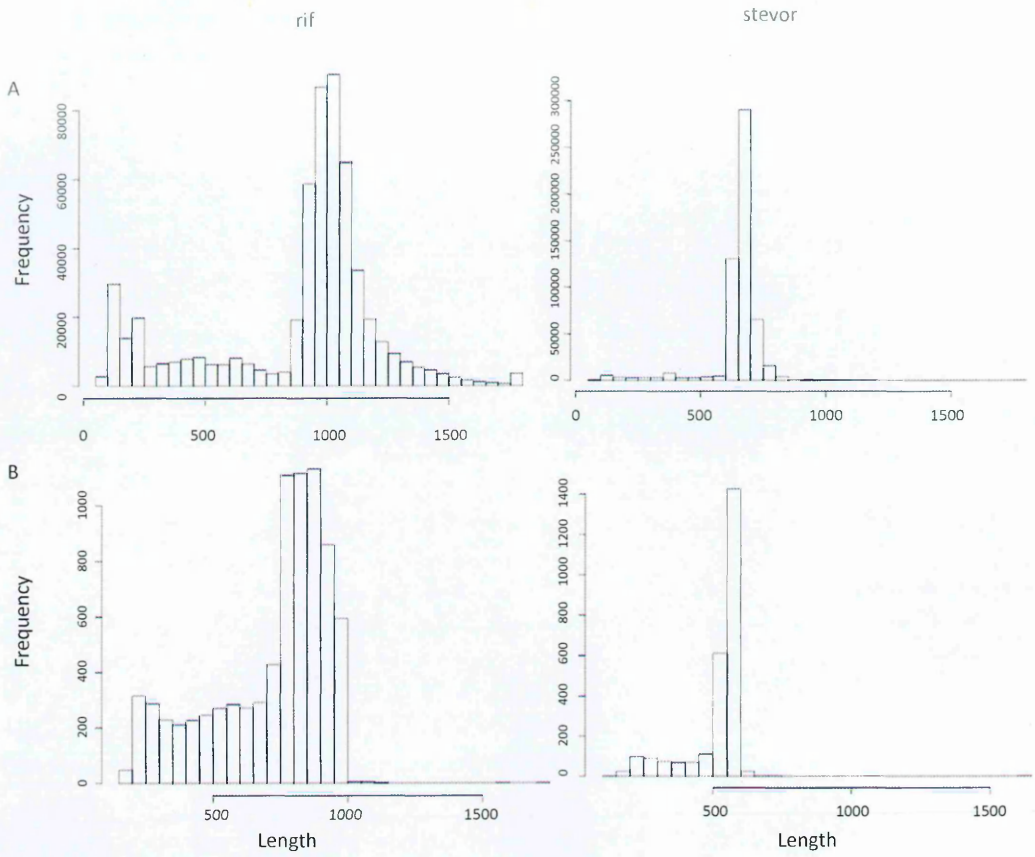


Figure 6.1. *Length distribution graphs for raw reads and assembled and cleaned 454 contigs from rif (left) and stevor (right). A) Raw 454 reads from all amplicons sequenced prior to assembly, 563,980 rif reads (left) and 793,118 stevor reads (right). B) Newbler assembled contigs after removal of non-target sequences from all the isolates, 7931 rif contigs (left) and 2625 stevor contigs (right).*

6.4.2. Cleaning up *rif* and *stevor* de novo assembled contigs.

Cleaning up the contigs by using BLAST was carried out to remove non-target sequences. A total of 8246 *rif* and 2601 *stevor* contigs were assembled from the 454 amplicon sequencing reads above. An initial BLAST analysis of all contigs against NCBI nucleotide nr database identified 80 non-target sequences from *rif* contigs and 75 non-target sequence from *stevor* contigs (Figures 6.1 and 6.2). The remaining 8166 rif and 2526 stevor contigs were searched against *rif* and *stevor* databases that contained sequences from laboratory and field isolates from Brazil, Gabon and Kenya, as well as amplicon sequences from Kenyan isolates generated by capillary sequencing (discussed in Chapter 4). For *stevor* all the remaining contigs had matches to stevor sequences in our database. For *rif* 108 contigs had no hit after BLAST searches for the remaining 8166 contigs against *rif* database. Because of the high diversity within the *rif* gene family it is possible that some divergent *rif* contigs had no match to any of the *rif* in our database. A better approach was used, where all the 8166 contigs were clustered using networks and a representative sequence from each cluster was searched by BLAST against the NCBI nucleotide nr database. Of these contigs 7955 hit against *rif* sequences, with the rest hitting sequences from other Plasmodium genes, human genes and bovine genes. The remaining 7955 contigs were clustered (discussed in subsequent section). MEGABLAST analysis was run for each cluster was against our *rif* database as well as NCBI nucleotide nr database. This searches identified an additional 24 non-rif contigs. The remaining 7931 contigs had matches to *rifs* in our database (Table 6.1).

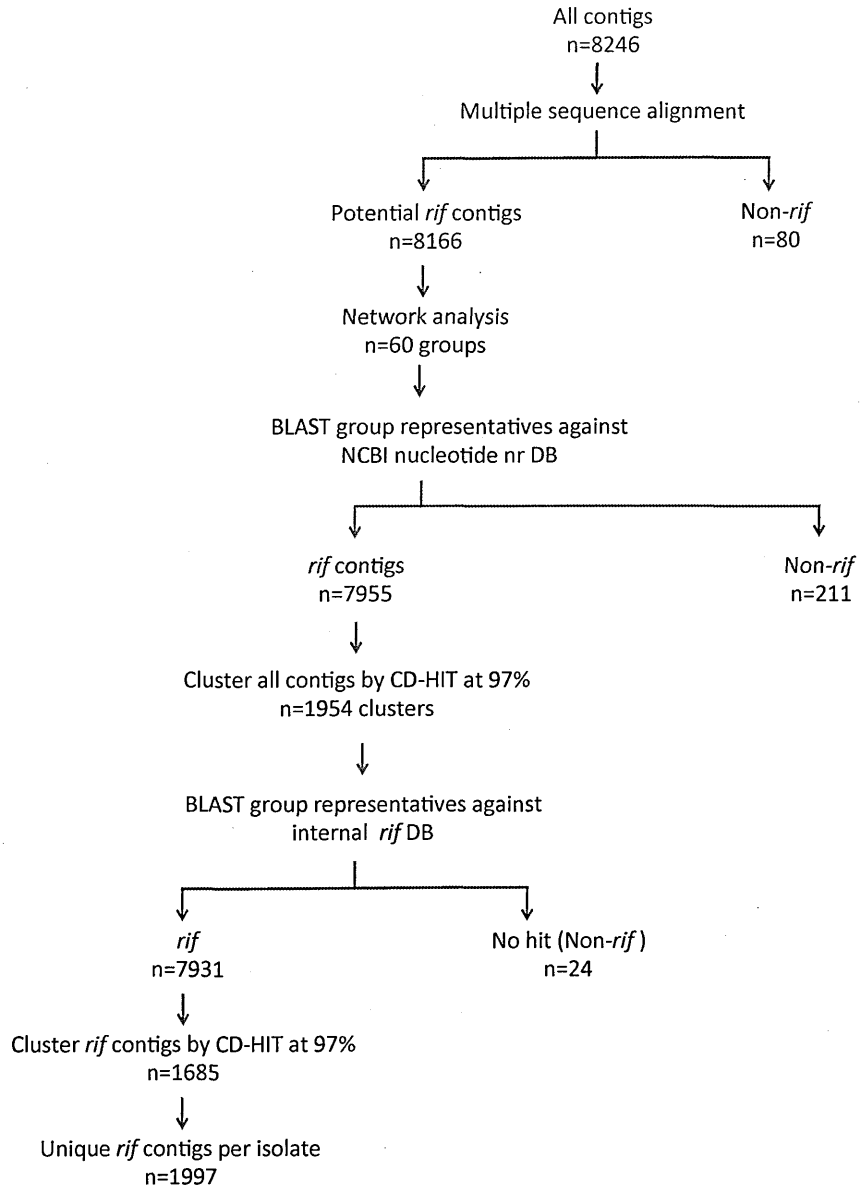


Figure 6.2. *Flow diagram of rif contig analysis for amplicons, from all the isolates, sequences by 454. 8246 contigs were obtained by assembling 454 reads from rif amplicons. After cleanup 7391 rif contigs were obtained. These contigs were clustered into 1685 unique rif variant types, using CD-HIT. The total number of unique variants from all the isolates was 1997.*

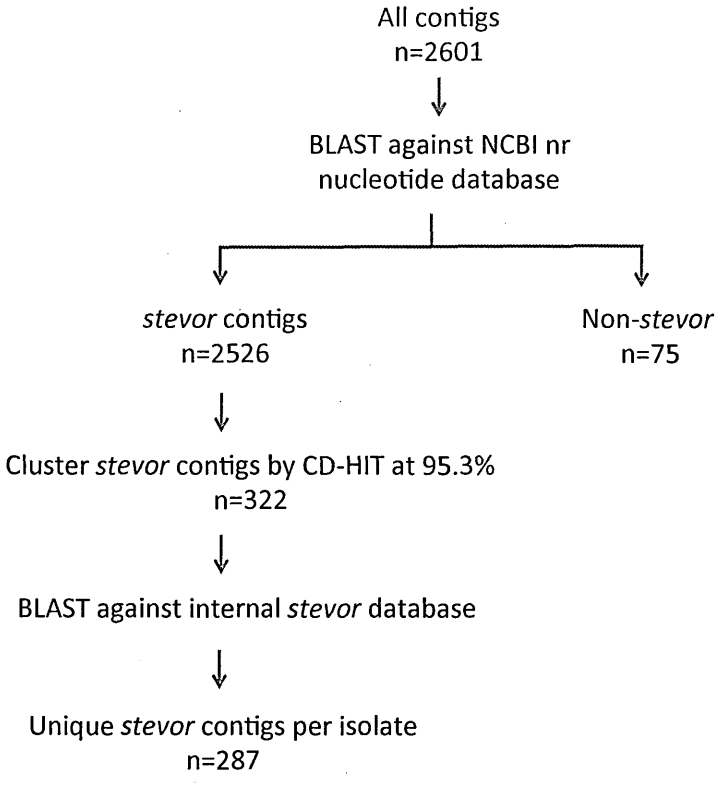


***Figure 6.3**. Flow diagram of stevor contig analysis. 2601 contigs were obtained by assembling 454 reads from stevor amplicons from all the isolates, sequences by 454. After cleanup 2526 stevor contigs were obtained. The contigs were clustered into 322 stevor variant types using CD-HIT. A second BLAST analysis against a database of all stevor from published and unpublished repertoires collapsed the number of unique variants from all isolates to 287.*

The non-target sequences from *rif* contigs were predominantly made up of the carrier DNA of bovine origin that was added to samples of low concentration prior to sequencing (Figure 6.4.A). The second most abundant contaminant was other *P. falciparum* genes including 18s and 28s ribosomal RNA (rRNA) that were non-specifically amplified by the primers. For stevor the predominant contamination was from human genomic material followed by other *Plasmodium* genes mostly ribosomal RNA (Figure 6.4.B). Two *var* gene sequences were recovered from contigs assembled from *stevor* amplicons.

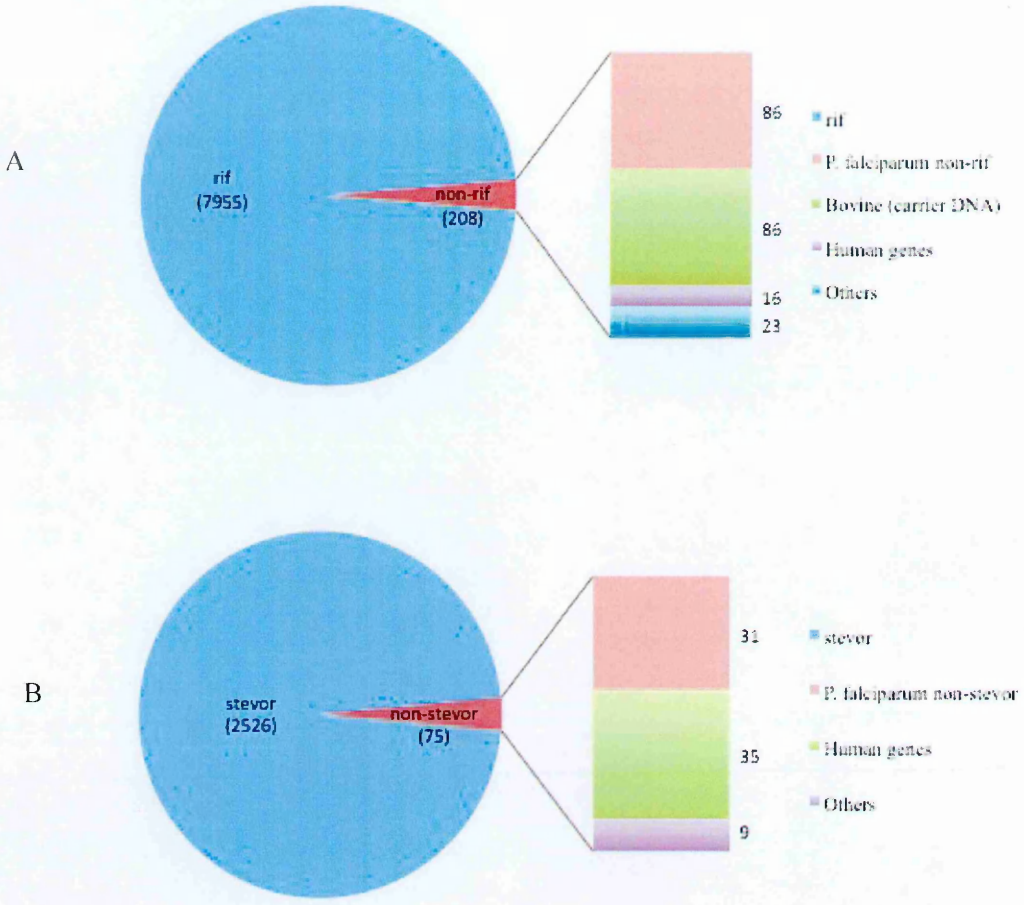


Figure 6.4. *Composition of contigs generated by de novo assembly of reads from rif (A) and stevor (B) amplicons from 3 laboratory and 11 wild isolates. pir sequences (blue region of the pie chart) were predominantly recovered but there was a small number of non-target contaminating sequences displayed in the bar graph.*

Table 6.1. *Summary of de novo assembly of 454 sequence reads from rif and stevor amplicons from 3D7 and wild isolates that lacked WGS data.*

Isolate	rif (563,980 reads)			stevor (793,118 reads)		
	Total contigs	rif contigs	Unique rif contigs	Total contigs	stevor contigs	Unique stevor contigs
3D7	532	508	126	180	180	29
8383	692	686	137	189	180	21
10727	382	362	193	166	163	44
10735	196	181	124	28	28	27
10739	173	155	137	30	30	30
10747	417	410	241	203	201	53
10761	340	319	206	43	43	27
10814	383	355	355	121	112	56

6.4.3. Clustering *rif* and *stevor* de novo assembled contigs into unique sequence types.

The assembly of reads into contigs was carried out per sample (i.e. per time point) as described earlier (Section 6.3.1). A set of contigs was therefore obtained from each time point within an isolate. In order to relate the contigs to one another across the whole time series dataset, and to account for abundant sequence errors, *rif* and *stevor* contigs were clustered by percent sequence identity resulting in unique contigs that represent different *pir* variants. This was important to enable temporal expression analysis of *pir* variants across time points as well as across isolates. The online CD-HIT application was used for sequence clustering, using cd-hit-est (Huang et al., 2010) (Figure 6.2 and 6.3). The application clusters sequences based on a set identity threshold (see section 6.3.3 above). The sequences are ordered by length and the longest sequence becomes the cluster representative. All sequences are compared to the representative and if they fail to meet the identity threshold they form the representative of the next cluster. In this way clusters, which represent unique gene variants, are generated with the longest sequence within the cluster as the cluster representative.

6.4.4. Determination of sequence identity thresholds using the 3D7 *rif/stevor* repertoire in order to derive unique variant types

In order to collapse all the *pir* contigs into unique clusters it was important to identify the optimal similarity threshold for the CD-HIT clustering algorithm. Because of the high error rates of 454 sequencing a very high clustering stringency resulted in non-unique *pir* clusters (Figure 6.5). 3D7 was used to optimize clustering stringency because both

mapping data from 454 amplicon sequencing as well as whole genome sequence data was available. Reads from 3D7 amplicon samples were assembled into 511 *rif* and 180 *stevor* contigs. The contigs were clustered using CD-HIT with identity thresholds between 80% and 100% (Figure 6.5). The clustering stringency was considered to be optimal when the number of clusters from the 454 de novo contigs was equal to the number of clusters from the unique *pir* variants observed in the mapping data. The mean percent pairwise identity cutoff was set at 97% for *rif* and 95.3% for *stevor*.

All the contigs were then clustered using the thresholds identified above initially resolving into 1954 unique *rif* from the 7955 contigs, and 322 unique *stevor* from the 2526 contigs (6.2 and 6.3). A limitation that the clustering approach was not able to resolve was the short read lengths obtained after 454 sequencing (see Chapter 5 Figure 5.3 and 6.3 above). The short reads assembled into short contigs much shorter than the expected length of *rif* and *stevor* amplicons (Chapter 5 Figure 5.1). Therefore some of the unique contigs identified might represent non-overlapping fragments of the same gene. For this reason a second BLAST search against our *rif* and *stevor* database was carried out to further cluster sequences into truly unique variants. The BLAST analysis enabled identification of 24 non-*rif* sequences from the 7955 rif contigs, which were discarded from the analysis. The remaining 7931 contigs were clustered at 97% identity cutoff resulting into 1685 unique clusters (Figure 6.2). Some of the unique *rif* genes were conserved across isolates and therefore the contigs were split up into their respective isolates of origin resulting in a total of 1997 unique *rif* clusters (Figure 6.2). For *stevor* contigs the 322 clusters were further collapsed into 287 unique clusters (Figure 6.3).

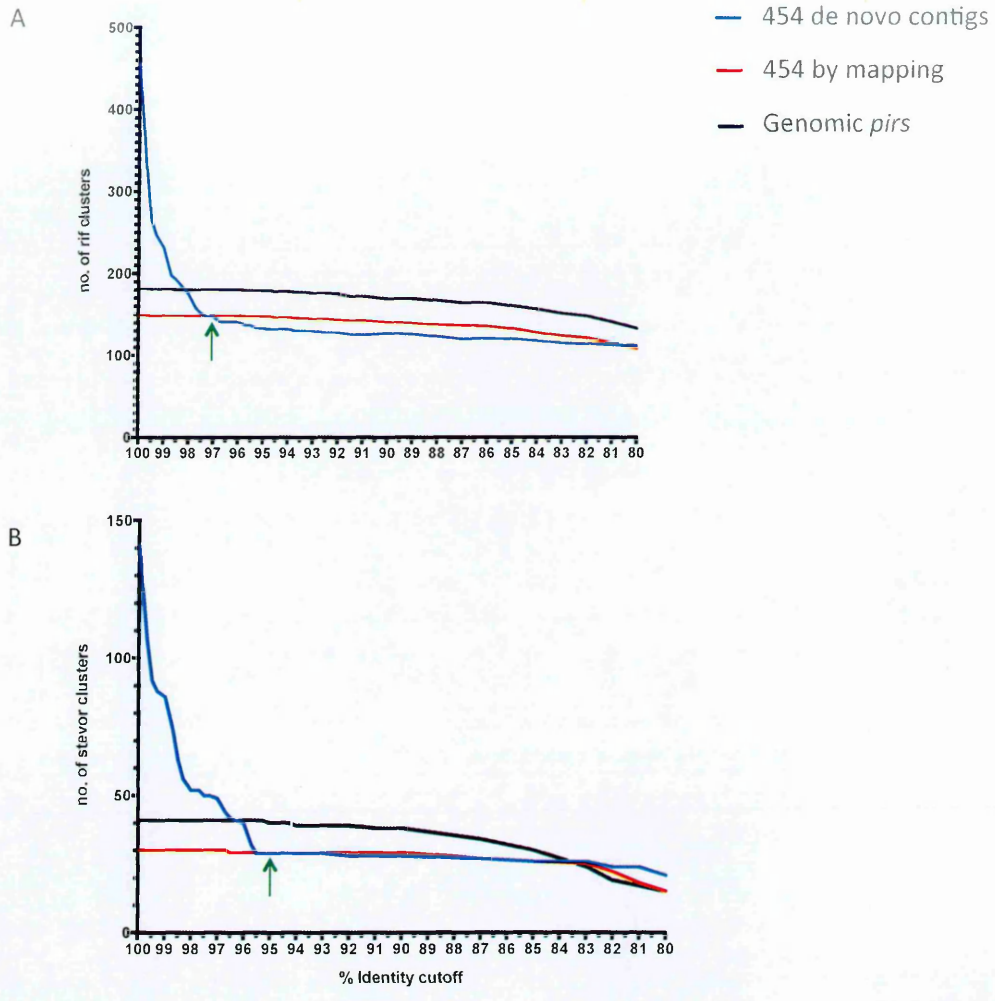


Figure 6.5. *Plot showing the number of 3D7 rif (A) and stevor (B) clusters obtained by running CD-HIT at different identity thresholds (80-100%), on 454 de novo assembled contigs (blue line), pirs recovered from 454 mapping data (red line) and genomic pir repertoires from 3D7 v3 genome (black line). The green arrows indicate the optimal clustering thresholds for rif (97%) and stevor (95.3%) where the blue line (de novo contigs) touched the red line (unique rif by from 454 mapping data).*

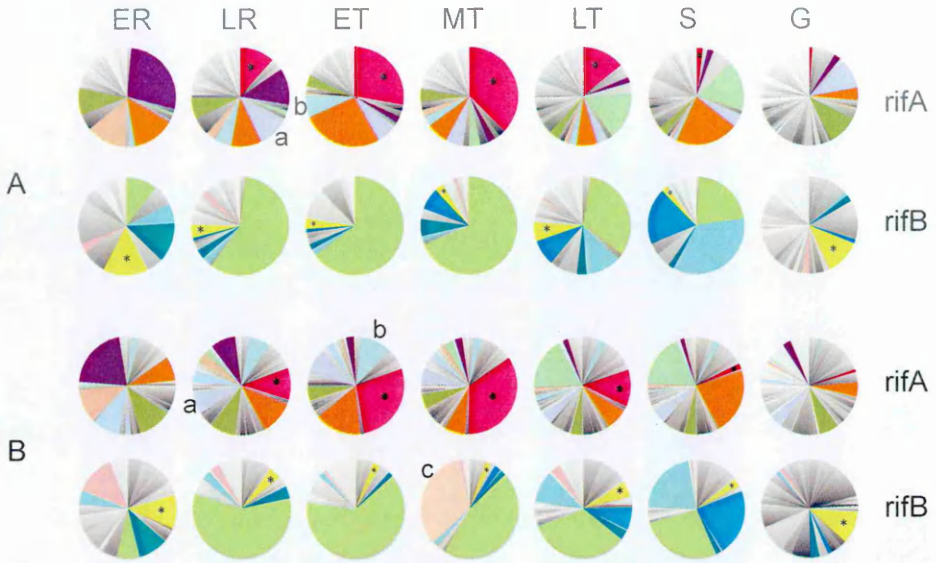
6.4.5. Validation of the de novo assembly analysis approach.

A comparison of the results obtained by de novo assembly of contigs and mapping reads onto the genome was carried out to determine whether similar repertoire coverage and transcript abundance were attained. The comparison was based on 3D7 amplicon sequence reads.

For both *rif* and *stevor* similar overall results were observed (Figure 6.6. I), with only slight differences in proportions of certain genes in some time points. For *rif* the similar dominant transcripts were observed, with the exception of a few, with high correlation between results obtained by the two methods (Figure 6.6. II). The most notable difference was in the mid trophozoite stage (MT) where the dominant *rifB* PF3D7_1372600 by mapping was not recovered in the de novo analysis data. One *rifA* transcript, PF3D7_011560 that predominated expression in the early trophozoites in the mapping data (11% transcript abundance) was not as highly expressed in the de novo data (7% transcript abundance). Another *rifA* PF3D7_110030 that was not highly expressed in the mapping data came ups as a dominant transcript in the de novo data (11% transcript abundance) at the late ring stage (Figure 6.6. I and III). For *stevor* transcript abundance across the time points was similar. The major difference was in the gDNA where the correlation between de novo and mapping data was much lower (0.387) than in the cDNA (Figure 6.7. II). This was mainly caused by the difference in recovery of one gene PF3D7_0101800. This stevor gene was abundant in the mapping data (9% of gDNA reads mapped) but was completely missed in the de novo assembly data. Additionally there was a significant difference in the number of reads assigned to one gene PF3D7_1479500 in the mapping data (6% of gDNA reads mapped) which was only

recovered at low abundance in the de novo data, with only 1% of the gDNA reads belonging to this gene (Figure 6.7.I and III).

I



II

3D7 rif	ER	LR	ET	MT	LT	S	G
Correlation between mapping and de novo data	**0.959**	**0.972**	**0.956**	**0.982**	**0.990**	**0.977**	**0.911**
Total reads by mapping	3877	6290	4407	3642	4702	4432	4904
Total reads by de novo assembly	2786	4465	3208	2294	3156	2917	3608
Number of unique variants by mapping	102	121	100	103	113	104	115
Number of unique variants by de novo assembly	48	79	47	54	74	62	101

III

Label	*rif/stevor*	Cluster ID	New ID	Old ID	*rif*A/B
	rif	*RIF_0193*	PF3D7_1300400	PF13_0004	A
	rif	*RIF_0326*	PF3D7_1150300	PF11_0520	A
	rif	*RIF_0280 (a)*	PF3D7_1100300	PF11_0009	A
	rif	*RIF_0336*	PF3D7_0701100	new rifA	A
	rif	*RIF_0284*	PF3D7_0425900	PFD1240w	A
	rif	*RIF_0001*	PF3D7_0421500	PFD1020c	A*
	rif	*RIF_0294 (b)*	PF3D7_0115600	PFA0760w	A
	rif	*RIF_0255*	PF3D7_1200500	PFL0025c	A
	rif	*RIF_0537*	PF3D7_1400300	PF14_0003	B
	rif	*RIF_0277*	PF3D7_1000300	PF10_0003	B
	rif	*RIF_0409*	PF3D7_0701200	new rifB	B*
	rif	*RIF_0043*	PF3D7_1000600	PF10_0006	B
	rif	*RIF_0276*	PF3D7_1300600	PF13_0006	B
	rif	*RIF_0395*	PF3D7_0900500	PFI0025c	B
	rif	*NA (c)*	PF3D7_1372600	MAL13P1.495	B

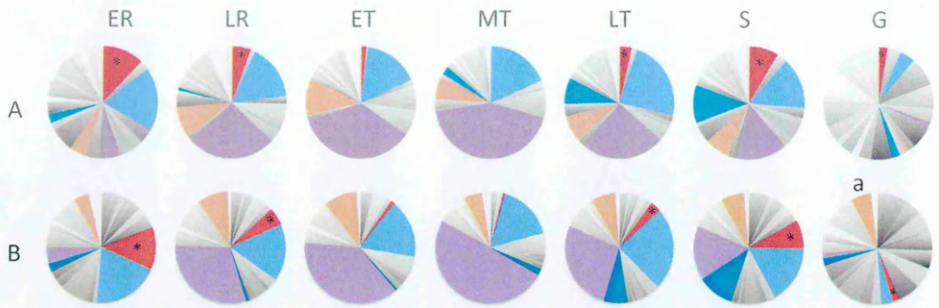
Figure 6.6. *(I) rif transcript abundance by (A) de novo assembly of reads into contigs (B) mapping of reads onto the genome. The pies represent proportions of reads that mapped to rifA (top panel) and rifB (lower panel Coloured regions of the cDNA pies represent transcripts that accounted for 10% and above of the total expression in at least one time point, while in gDNA pies, coloured regions indicate genes whose read counts represent >10% of the total reads obtained from amplification of matched cDNA. (II) Correlation between rif transcript abundance by de novo contig assembly and mapping. (III) Table showing most abundant rif transcripts by both analysis methods.*

a - PF3D7_1100300, a dominant rifA transcript in LR in de novo data (11% of the proportion) but not dominant by mapping data (9%) .

b – PF3D7_0115600, a dominant rifA transcript in ET in mapping data (11% proportion) but not dominant in the de novo data (7%).

c – PF3D7_1372600 a dominant rifB transcript in MT in the mapping data (33% proportion) but completely missing in the de novo data.

I



II

Time point	ER	LR	ET	MT	LT	S	G
corelation between mapping and de novo data	**0.998**	**0.995**	**0.997**	**0.995**	**0.998**	**0.997**	**0.387**
Total reads by mapping	2340	4714	2212	3046	2873	4093	2231
Total reads by de novo assembly	2032	3914	1783	2372	2466	3509	1495
Number of unique variants by mapping	28	28	29	27	28	27	29
Number of unique variants by de novo assembly	27	27	19	20	25	27	28

III

Label	Cluster ID	New ID	Old ID
	STE_454_006	PF3D7_0617600	PFF0850c
	STE_454_001	PF3D7_0401500	PFD0065w
	STE_454_038	PF3D7_1040200	PF10_0395
	STE_454_092	PF3D7_0901600	PFI0080w
	STE_454_082	PF3D7_1479500	PF14_0767

Figure 6.7. *(I) stevor transcript abundance by (A) de novo assembly of reads into contigs (B) mapping reads onto the genome. a: (II) Correlation between stevor transcripts abundance by de novo contig assembly and mapping reads to the genome. (III) Table showing most abundant stevor transcripts that were similarly recovered by both analysis approaches.*

a: stevor PF3D7_1479500 predominantly recovered in 3D7 gDNA amplicon sequence data by mapping but very few reads by de novo assembly of contigs.

6.4.6. Analysis of *pir* expression profiles in patient isolates.

To analyse *rif* and *stevor* expression profiles in the patient isolates unique *pir* contigs from the de novo assembly data were considered as unique variants. For each isolate a set of unique variants (unique contigs) were generated. Transcript abundance for each unique variant was determined from the number of reads that were used in generation of the contigs. For contigs that were collapsed by CD-HIT clustering or by BLAST analysis the number of reads that generated the contigs were summed up.

For many of the patient isolates the amount and quality of RNA and DNA was low, therefore amplifying *pir* genes from this material was problematic, and completely failed for some samples. In addition the significant setback encountered during 454 sequencing where two sequencing runs failed and the successful run resulting in suboptimal read lengths subsequently lead to poor sequence data (discussed in Chapter 5). Due to these limitations the data on *pir* repertoires and transcription patterns described in subsequent sections was incomplete for some patient isolates (Table 6.2). For the samples with adequate data transcript abundance was analysed for each contig, the contigs representing unique *pir* variants. Transcript abundance was considered the proportions of reads that went into assembly of a contig out of the total number of reads of all the contigs in a time point.

Table 6.2. *Summary of the 454 de novo assembled contigs data from patient parasite isolates.*

Isolate	Clones by MSP2 genotyping	IDC stage	rif			stevor		
			Total reads	Reads assembled into contigs	Unique rif contigs	Total reads	Reads assembled into contigs	Unique stevor contigs
8383 Acute	1	R	1425	973	48	5510	4304	17
		T	2296	2110	83	18573	17857	20
		Genomic	2443	1804	96	11399	9233	19
8383 Culture-adapted	1	R	2459	1884	48	x	x	x
		LR/ET	2614	1684	46	7,334	5898	16
		T	2877	1776	55	7,869	5141	13
		S	3819	2558	64	6,077	3914	18
		Genomic	2352	1596	98	2268	1686	17
10727	1	ER	ns	ns	ns	782	465	14
		R	1257	308*	14	7790	6890	29
		T	3186	2108	77	6888	7406	29
		S	1328	40*	6	6390	5562	30
		Genomic	3424	3269	170	6512	5668	29
10735	nd	MRI	4238	258*	32	ns	ns	ns
		LR/ET	443	155*	9	ns	ns	ns
		ET	2489	448*	27	ns	ns	ns
		Genomic	3808	2406	109	3517	2957	27
10739	4	R	3080	56*	7	ns	ns	ns
		T	1875	162*	15	ns	ns	ns
		Genomic	7129	5023	129	6725	6464	30
10747	4	RN	4153	2671	111	6103	5509	46
		ET	3413	2045	134	5056	4343	47
		ND-60	2481	1463	105	6190	5278	46
		Genomic	8694	301*	35	7492	6589	49
10761	1	MR	2603	653	41	ns	ns	ns
		ET	ns	ns	ns	2853	2125	13
		LT	4464	32*	5	ns	ns	ns
		S-ER2	3247	2873	123	1391	1111	26
		Genomic	2645	1683	125	ns	ns	ns
10814	2	ER	6664	2043	41	ns	ns	ns
		RN	1993	185*	21	2579	1473	38
		MR	2407	211*	17	ns	ns	ns
		LR	1580	52*	4	ns	ns	ns
		PR	3745	1983	46	2108	1548	21
		ET	2803	513	41	ns	ns	ns
		Genomic	4597	3221	176	4554	4131	50

ER: early rings

R/ RN: rings

MR: mid rings

LR: late rings

ET: early trophozoites

PR: late ring to early trophozoites

T: trophozoites

S: schizonts

**: Sub-optimal read depth (<500 reads)*

ns: no sample (cDNA or gDNA sample or amplicon)

x: sequencing failed

6.4.7. Differential expression of *pir* genes in a patient isolate before and after culture adaptation.

A field isolate 8383 sampled from a patient with non-severe malaria that had been culture adapted was obtained from an ongoing *var* study (led by Dr. Pete Bull). Expression profiling for *rif* and *stevor* genes was carried out in the ex vivo ring and trophozoite samples, and at four time points across the IDC in the culture-adapted parasite line to test the effects of different selection pressures on *pir* gene expression. There was a clear difference between ex vivo parasites and culture adapted parasites in *rifA* expression. Four dominant transcripts were observed in rings, one of which was shared between the acute and cultured parasite lines (Figure 6.8). The dominant *rifA* transcript in ex vivo rings 454_RIF_0002 is conserved in IT, DD2, HB3 and 5 Kenyan isolates. The ortholog of this *rif* PF8443_8_7_U439800 was also predominantly expressed in ex vivo rings of a severe malaria isolate 9106 (discussed in Chapter 5). The dominant *rifA* transcript cultured rings 454_RIF_0080 remained dominant in cultured late ring and early trophozoites and schizonts. In ex vivo trophozoites no clearly dominant *rifA* was detected but in the cultured trophozoites the conserved pseudogene 454_RIF_0001 (PF3D7_0412600 in 3D7) predominated expression. Similar dominant *rifB* transcripts were observed in acute and cultured parasites, with the exception of PF3D7_0711900 that was only seen in the cultured rings, and 454_RIF_0286 mainly recovered in the cultured schizonts (Figure 6.8). For stevor a very strict stage-specific expression profile was observed. There were clearly different dominant transcripts at the different.stages. 5 dominant *stevor* transcripts were detected in acute rings, one of which was a conserved stevor pseudogene 454_STE_001 (PF3D7_0401500 and PFIT_0400200 in 3D7 and IT respectively), which was predominantly expressed in 3D7 (Figure 6.7) and IT (Figure

5.12) rings, and in HB3 in the mid and late stages (Figure 5.14). This pseudogene dominated stevor expression in both the acute and cultured trophozoites. 3 new dominant transcripts were observed in the cultures schizonts. For *stevor* there was a bias towards the gene STE_454_103 based on the number of gDNA reads assigned to this gene (Figure 6.8). This gene came up as a dominant ring transcript in the acute parasite but this could have been as a result of the bias.

8383 NS	Ring cDNA			LR / ET cDNA			Trophozoite cDNA			Schizont cDNA			Genomic DNA		
	rifA	*rifB*	*stev*	*rifA*	*rifB*	*stev*	*rifA*	*rifB*	*stev*	*rifA*	*rifB*	*stev*	*rifA*	*rifB*	*stev*
															
															

Cluster ID	pir	Type	Cluster representative	BLAST hit	Identities
454_RIF_0002	*rif*	A	PF383_RIF_AC_00_R1_2_14	PFDG_00501.1	886/890 (99%)
454_RIF_0001*	*rif** (c)	A	PF383_RIF_CU_00_G2_4	PF8443_8_4_U067000*	850/862 (98%)
454_RIF_0080	*rif*	A	PF383_RIF_CU_00_G1_2_3	PfHB3_BroadWash527800	788/985 (80%)
454_RIF_0120	*rif*	A	PF383_RIF_CU_00_G1_2_47	PF8443_8_4_U517400	823/856 (96%)
454_RIF_0177	*rif*	A	PF383_RIF_AC_00_GE_2_19	PF3D7_1000200 (PF10_0002)	766/953 (80%)
454_RIF_0171	*rif*	BX	PF383_RIF_AC_24_T1_2_42	PF8443_8_3_U561800	705/728 (96%)
454_RIF_0012	*rif*	B	PF383_RIF_AC_24_T1_2_15	PF3D7_0833200 (MAL7P1.215)	891/899 (99%)
454_RIF_0116	*rif*	B	PF383_RIF_AC_00_GE_2_31	PF8443_8_4_U510600	719/873 (82%)
454_RIF_0143	*rif*	B	PF383_RIF_CU_30_SZ_2_13	DQ265331	868/905 (95%)
454_RIF_0275	*rif*	B	PF383_RIF_AC_24_T2_2_31	PF8443_8_8_U501800	676/779 (86%)
454_RIF_0286	*rif*	B	PF383_RIF_AC_24_T1_2_39	PF8443_8_4_U507500	719/756 (95%)
STE_454_001*	*stevor** (c)	na	PF383_STE_CU_06_P1_4	PF3D7_040150 (PFD0065w)*	592/595 (99%)
STE_454_033	*stevor*	na	PF383_STE_CU_00_GE_3_2	DQ265689	491/564 (87%)
STE_454_050	*stevor*	na	PF383_STE_CU_18_TR_1	PF3D7_0900900 (PF10045c)	523/558 (93%)
STE_454_051	*stevor*	na	PF383_STE_CU_30_SZ_7	PF3D7_1372800 (MAL13P1.505)	548/582 (94%)
STE_454_057	*stevor*	na	PF383_STE_AC_00_R1_3_9	PFSAO75_PN64	459/572 (80%)
STE_454_078	*stevor*	na	PF383_STE_AC_00_G2_3_6	PF8443_8_3_U499400	561/562 (99%)
STE_454_089	*stevor*	na	PF383_STE_AC_24_T1_10	PF8443_8_1_U516500	509/535 (95%)
STE_454_103	*stevor*	na	PF383_STE_CU_00_GE_3_9	PF8443_8_4_U542800	537/556 (96%)
STE_454_115	*stevor*	na	PF383_STE_AC_00_R1_3_10	PF8443_8_3_U446600	557/589 (94%)
STE_454_153	*stevor* (c)	na	PF383_STE_CU_30_SZ_10	PF3D7_0832400 (MAL7P1.223)	565/569 (99%)
STE_454_002*	*stevor** (c)	na	PF383_STE_AC_00_R1_3_1	PFDG_03937*	595/598 (99%)

Figure 6.8. rif and stevor genomic repertoires and transcription profiling in an ac and culture-adapted patient isolate 8383 at ex vivo rings trophozoites (top pane after culture adaptation at 4 time points rings, late ring to early trophozoites (L trophozoites and schizonts. (A) Piecharts representing proportions of reads that up the rifA (top panel), rifB (middle panel) or stevor (lower panel) de novo as contigs. Coloured regions of the cDNA pies represent transcripts that account 10% and above of the total expression in at least one time point, while in gDNA coloured regions indicate genes whose read counts represent >10% of the tot obtained from amplification of matched cDNA (see table below). (B) Table showi the dominant transcripts (>10% of total transcript abundance) at any time poi the BLAST search results against rif/ stevor databases.

6.4.8. *Rif* and *stevor* transcript upregulation in a patient isolate (8383).

Analysis of *pir* differential expression was carried out using DESEQ in R (described in Chapter 5). Upregulation was defined as the number of reads from cDNA divided by the number of reads from gDNA.

The temporal expression pattern of *rif* genes could be seen from transcript upregulation analysis (Figure 6.9. A). *rif* genes in cluster 1 were upregulated at all stages in the acute and culture-adapted parasite lines. Cluster 2 *rif* genes were mostly upregulated in the cultured parasite line, at ring, trophozoite and schizont stages. Cluster 3 was mostly up in trophozoite stages in the acute and cultured parasites, although a subset was upregulated in acute rings. Cluster 4 was upregulated in the early stages, in acute rings and cultured rings and late ring/ early trophozoite stage. Cluster 5 stayed low through out the IDCs of acute and cultured parasites.

There was good correlation between transcript abundance and transcript upregulation for *stevor* from 8383 (Figure 6.9.B). *Stevor* clustered into 3 groups based on transcript upregulation. Genes in cluster 1 were upregulated in all time points but peaked in trophozoites of the cultured parasites. This cluster was made up of the dominant stevor transcripts (Figure 6.8). Cluster 2 was upregulated in schizonts of the cultured line while cluster 3 was not upregulated at any stage.

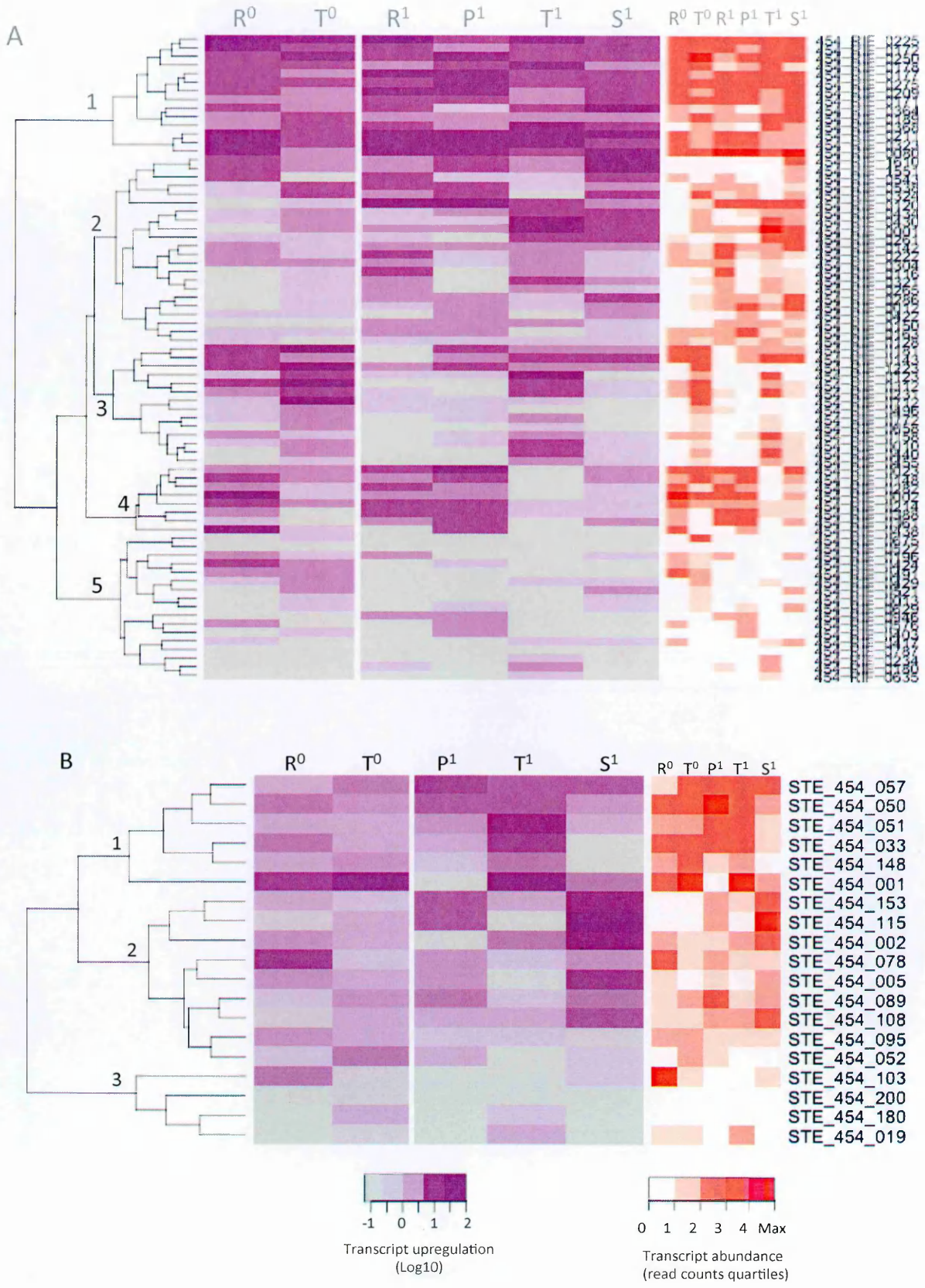

A
R0 T0 R1 P1 T1 S1
R0 T0 R1 P1 T1 S1
1
2
3
4
5
B
R0 T0 P1 T1 S1
R0 T0 P1 T1 S1
1
2
3
STE_454_057
STE_454_050
STE_454_051
STE_454_033
STE_454_148
STE_454_001
STE_454_153
STE_454_115
STE_454_002
STE_454_078
STE_454_005
STE_454_089
STE_454_108
STE_454_095
STE_454_052
STE_454_103
STE_454_200
STE_454_180
STE_454_019
-1 0 1 2
Transcript upregulation
(Log10)
0 1 2 3 4 Max
Transcript abundance
(read counts quartiles)

Figure 6.9. *Transcript upregulation of pir genes in acute and culture-adapted Kilifi patient isolate 8383. Heatmaps showing rif (A) and stevor (B) transcript upregulation (purple) and transcript abundance (orange) in ex vivo rings (R^0) and trophozoites (T^0) and in cultured parasites at 4 IDC stages, ring (R^1), late ring to early trophozoites (P^1), trophozoite (T^1), and schizont (S^1). The left panel (purple) indicates log10 fold change transcript upregulation across the IDC. The right panel (orange) represents transcript abundance categorized into quartiles. The dendogram was generated by hierarchical clustering in R based on transcript upregulation pattern.*

6.4.9. *Pir* expression profiling in patient isolates.

Expression patterns of *rif* and *stevor* genes were analysed in 6 ex vivo parasite isolates through a single IDC. The ex vivo time point samples had been collected for an ongoing microarray study (led by Dr. Margaret Mackinnon) as previously discussed (Chapter 5). For many of the patient isolates analysed in this section only small amounts of RNA were available for the *pir* expression study. For this reason there were gaps in the *pir* genomic and expression data in the patient isolates, a major limitation in the analysis as described above (see table 6.2). For some samples there was adequate data to analyse *pir* expression across time points (Figure 6.10 and Table 6.3). Transcript abundance was analysed from the available data but transcript upregulation could not be analysed because of the missing time points as well as few number of reads from the available time points and gDNA samples.

As previously observed *rif* expression varied across IDC stage in an isolate specific manner (Figure 6.10). For the isolate 10727 there was limited data for rings and schizonts (308 and 40 reads respectively). A single dominant *rifA* transcript 454_RIF_0009 was observed in rings. Although this could be as a result of the missing data, expression of an ortholog in rings stages of a different patient isolate supports this observation. The ortholog, a pseudogene, PFSAO75_PN48, (99% pairwise identity) was also dominantly expressed in rings and schizonts of the rosetting patient isolate SA075 (discussed Chapter 5). This *rif* was closely related to the 3D7 PF3D7_1372700 that was not expressed in 3D7 time points. This *rif* went off in the trophozoites and three new dominant transcripts came up, one of which was carried on into schizonts. Because of the limited number of reads in schizonts (40 reads) it was not possible to determine the dominant schizont transcripts (Figure

6.10). Different *rifB* transcripts were detected at the different stages, one shared between rings and trophozoites. Only two *rifB* variants were detected in schizonts. For *stevor* 4 transcripts were highly expressed in 10727 (Figure 6.10). The 2 dominant ring transcripts observed stayed on in later stages but at low levels. Two transcripts dominated expression in the late stages, including the conserved pseudogene STE_454_001 that was dominant in the patient isolate 8383 discussed above (Figure 6.8).

For the isolate 10747, a large number of *pir* variants was detected, 241 *rif* and 53 *stevor* (Table 6.1) indicating more than one repertoire due to multiple parasite genomes in this isolate (4 MSP2 genotypes detected, Table 6.2). For this reason interpreting the expression data was complicated because of the many transcripts detected (Figure 6.10). *rifA* was similar at all time points, dominated by two transcripts. *rifB* expression changed only slightly between rings and late ring/ early trophozoite stage. 4 dominant transcripts were observed in rings, one of these staying on as the dominant transcript in the following two stages. The late ring/ early trophozoite and schizont stage were similar in *rifB* expression profile. Similarly *stevor* expression profile changed very little across time points with only two transcripts dominating, STE_454_001 and STE_454_203.

10761 *pir* sequence data had many gaps, but the time-specific expression of *pirs* was observed (Figure 6.10). For *rifA* 3 dominant transcripts were detected in rings, one of them, 454_RIF_0298 was the only transcript detected in trophozoites and another, 454_RIF_0451, dominated expression in schizonts. For *rifB* 4 dominant transcripts were detected across the IDC. 2 of them, 454_RIF_0043 and 454_RIF_0673 were

shared in ring and trophozoite stages. One of the ring transcripts 454_RIF_0280 was also detected in schizonts. A different transcript 454_RIF_0060 was dominant in schizonts. *rif* gene expression for this isolate had been previously analyzed by cloning and capillary sequencing (Chapter 4). The dominant transcripts detected by capillary sequencing were also recovered in the 454 data with the exception of one (Figure 4.7), and at similar time points with slight differences in proportions. Some of the dominant *rif* transcripts in the capillary data were only detected in low amounts in the 454 data. The dominant ring and schizont *rifA* transcript in the capillary data (RIF_K011) was also dominant in rings and schizonts in the 454 data (454_RIF_0451). The dominant *rifA* transcript in trophozoites the 454 data (454_RIF_0298) was only a minor transcript in the capillary data in the early stages (RIF_K525). One of the dominant *rifB* in rings and schizonts in the 454 data was also dominant in schizonts in the capillary sequence data. One of the dominant *rifB* in schizonts in the 454 data (454_RIF_0060) was not captured in the capillary sequencing data.

Stevor expression profiling was possible for only two stages, the late ring/ early trophozoites and schizonts (Figure 6.10). For both stages the same 2 *stevor* genes were predominantly expressed, the conserved pseudogene STE_454_001 and STE_454_152. There was little similarity between the capillary sequence data and the 454 data. Only 8 stevor sequences were shared in the results of the two analyses, none of which was dominantly expressed at any stage.

For the isolate 10814 *pir* gene expression data was available for the first 2 stages of the IDC. For *rif* expression data from the early ring, the late ring/ early trophozoite and early trophozoite stages (ER, PR and ET in Table 6.2) was analyse for transcript abundance. For *stevor* data was available for only 2 stages, rings and late ring/ early

trophozoite stage (RN and PR in Table 6.2). The *pir* repertoire was larger than that of a single genome with 355 unique *rif* variants and 56 unique *stevor* variants (the expected repertoire is about 180 *rif* and 40 *stevor* genes) because of the multiple clones in this isolate as MSP2 genotyping detected 2 clones (Table 6.2). Despite the many *pir* variants detected in this isolate clear stage-specific dominant transcripts were detected (Figure 6.10). For *rifA* 2 dominant transcripts were observed in rings and a different set of 3 transcripts observed in the late ring and early trophozoite stage. For *rifB* unique transcripts were detected at each different IDC stage. Only one transcript was recovered in the ring stage. A different transcript was dominantly expressed in the late ring early trophozoites and a completely different set expressed in the early trophozoite stage. For this isolate the capillary sequence data and the 454 data showed some similarities. For the 2 stages that were analysed by both approaches (rings and late rings) two of the dominant *rifA* transcripts were shared in the 2 data sets. The ring transcripts 454_RIF_0580 and 454_RIF_0352 in the 454 data were also dominant in rings in the capillary data (RIF_K036 and RIF_K018 respectively). The late ring/ early trophozoite dominant *rifA* transcript 454_RIF_0397 and 454_RIF_0387 were dominant in the late rings in capillary sequence data (RIF_K058 and RIF_K027). A *rifB* late ring/ early trophozoite dominant transcript 454_RIF_0562 was predominantly recovered in trophozoites in the capillary data (RIF_K009).

The expression data for *stevor* in 10814 was only available for 2 stages, rings and late ring early trophozoite stage (Figure 6.10). In both stages the same two dominant stevor transcripts were observed, STE_454_057 and the conserved pseudogene STE_454_001 (Figure 6.10). The *stevor* gene STE_454_057 was also dominantly expressed in late rings/ early trophozoites in another patient isolate 8383 (Figure 6.8).

These two stevor genes were not detected in the capillary sequencing data (Chapter 4).

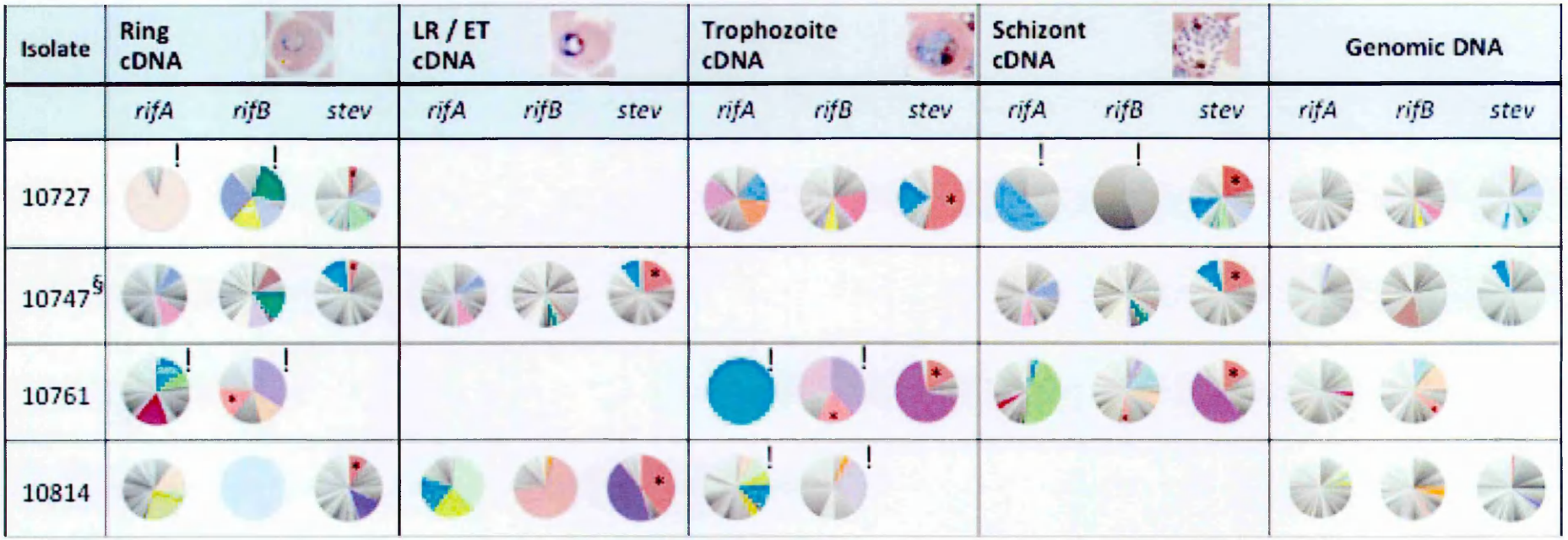


Figure 6.10. Expression profiling of pir genes in patient isolates. Piecharts showing rif and stevor genomic repertoires and transcript abundance in the IDC of patient parasite isolates, 10727, 10747, 10761 and 10814. Coloured regions of the cDNA pies represent transcripts that accounted for 10% and above of the total expression in at least one time point, while in gDNA pies, coloured regions indicate genes whose read counts represent >10% of the total reads obtained from amplification of matched cDNA (see Table 6.3 below). !: Poor sequence depth (below 1000 reads). §: Schizont stage in this parasite was not confirmed by slide or microarray data but estimated by time in culture (60 hours).

Table 6.3. Dominant rif and stevor transcripts in the IDCs of patient isolates

Cluster ID	pir	Type	Cluster representative	Isolate	Stage of peak expression	Capillary sequence ID	BLAST hit	Identities
454_RIF_0009	*rif* (c)	A	PF727_RIF_AC_00_ER_2_1	10727	R	N/A	PFSAO75_PN48	869/877 (99%)
454_RIF_0660	*rif*	A	PF727_RIF_AC_22_T2_2_29	10727	T	N/A	PF3D7_0700200 (MAL8P1.219)	716/783 (91%)
454_RIF_0481	*rif*	A	PF727_RIF_AC_00_G1_2_48	10727	T	N/A	PF8443_8_7_U558000	711/806 (88%)
454_RIF_0475	*rif*	A	PF727_RIF_AC_22_T1_2_21	10727	S	N/A	PF8443_8_7_U561000	838/844 (99%)
454_RIF_0237	*rif*	B	PF727_RIF_AC_00_G2_30	10727	R	N/A	PFIT_0100400	785/791 (99%)
454_RIF_0518	*rif* (c)	B	PF727_RIF_AC_22_T2_2_43	10727	R	735_g43	PF3D7_0632300 (PFF1570w)	595/599 (99%)
454_RIF_0564	*rif*	B	PF727_RIF_AC_00_G1_2_60	10727	R	N/A	PFDG_00191	735/778 (94%)
454_RIF_0621	*rif*	B	PF727_RIF_AC_22_T2_2_15	10727	R	N/A	PF8443_8_1_U512800	753/861 (87%)
454_RIF_0487	*rif*	B	PF727_RIF_AC_00_G1_2_69	10727	T	N/A	PfHB3_BroadWash516500	754/756 (99%)
454_RIF_0714	*rif*	A	PF747_RIF_AC_20_RN_2_7	10747	R	Srif19A	PF3D7_0632000 (PFF1555w)	856/918 (93%)
454_RIF_0479	*rif*	A	PF747_RIF_AC_40_ET_2_37	10747	S	N/A	PFDG_02074	707/812 (87%)
454_RIF_0559	*rif*	B	PF747_RIF_AC_00_GE_9	10747	R	594g_13	PF3D7_0222500 (PFB1005w)	696/736 (94%)
454_RIF_0605	*rif*	B	PF747_RIF_AC_60_ND_2_8	10747	R	10594-23	PF3D7_0732700 (PF07_0136)	825/880 (93%)
454_RIF_0662	*rif*	B	PF747_RIF_AC_60_ND_2_30	10747	R	N/A	PF8443_8_6_U546300	638/774 (82%)
454_rif_0669	*rif*	B	PF747_RIF_AC_60_ND_2_35	10747	S	N/A	PF8443_8_6_U544300	683/776 (88%)
454_RIF_0597	*rif*	A	PF761_RIF_AC_00_GE_2_11	10761	R	RIF_K071	PF3D7_0937700 (PFI1825w)	742/950 (78%)
454_RIF_0298	*rif*	A	PF761_RIF_AC_34_SZ_2_7	10761	T	RIF_K524	PF3D7_0200600 (PFB0035c)	760/944 (80%)
454_RIF_0451	*rif* (c)	A	PF761_RIF_AC_00_ER_2	10761	S	RIF_K011	DQ265365 (PFL2585c)	890/895 (99%)
454_RIF_0043	*rif* (c)	B	PF761_RIF_AC_16_MR_11	10761	R/T	RIF_K169	PF3D7_1000600 (PF10_0006)	839/854 (98%)
454_RIF_0673	*rif* (c)	B*	PF761_RIF_AC_34_SZ_2_60	10761	R/T	RIF_K032	PF3D7_0701200*	743/744 (99%)
454_RIF_0280	*rif*	B	PF761_RIF_AC_34_SZ_2_32	10761	R	RIF_K254	PF3D7_0900400 (PFI0020w)	814/853 (95%)
454_RIF_1030	*rif*	B	PF761_RIF_AC_26_LT_6	10761	T	RIF_K239	PF8443_8_6_U562600	649/661 (98%)
454_RIF_0060	*rif*	B	PF761_RIF_AC_34_SZ_2_59	10761	S	N/A	PFSAO75_101100	755/761 (99%)
454_RIF_0352	*rif*	A	PF814_RIF_AC_10_ER_1	10814	R	RIF_K018	PFSAO75_493700	764/951 (80%)
454_RIF_0580	*rif*	A	PF814_RIF_AC_10_ER_2	10814	R	RIF_K036	Gb_rif57	737/934 (78%)
454_RIF_0387	*rif*	A	PF814_RIF_AC_16_MR_5	10814	LR/ET	RIF_K027	PF8443_8_3_U532500	747/888 (84%)
454_RIF_0397	*rif*	A	PF814_RIF_AC_00_GE_2_57	10814	LR/ET	RIF_K058	PF3D7_0223100	727/845 (86%)
454_RIF_0442	*rif*	A	PF814_RIF_AC_16_MR_3	10814	LR/ET	RIF_K314	PF8443_8_10_U509700	701/791 (88%)
454_RIF_0317	*rif*	B	PF814_RIF_AC_10_ER_22	10814	R	N/A	PFSAO75_490000	759/771 (98%)
454_RIF_0562	*rif*	B	PF814_RIF_AC_16_MR_8	10814	LR/ET	RIF_K009	PFDG_04789	660/755 (87%)
454_RIF_0728	*rif*	B	PF814_RIF_AC_00_GE_2_51	10814	T	N/A	PF8443_8_1_U507700	850/857 (99%)
STE_454_073	*stevor*	na	PF727_STE_AC_22_T1_3_23	10727	R	nd	PFIT_bin09500*	530/544 (97%)
STE_454_109	*stevor*	na	PF727_STE_AC_22_T1_3_16	10727	R	nd	PfHB3_BroadWash162000	537/538 (99%)
STE_454_001*	*stevor* (c)	na	PF727_STE_AC_32_S1_3_9	all	T	nd	PF3D7_0401500 (PFD0065w)	590/594 (99%)
STE_454_156	*stevor*	na	PF727_STE_AC_22_T1_3_8	10727	T	nd	DQ265628.1|Pf1544_38	553/555 (99%)
STE_454_203	*stevor* (c)	na	PF747_STE_AC_00_GE_3_18	10747	R/S	nd	PF3D7_0300400 (PFC0025c)	496/538 (92%)
STE_454_152	*stevor*	na	PF761_STE_AC_34_SZ_3_4	10761	T/S	none	PF8443_8_1_U539200	568/575 (98%)
STE_454_057	*stevor*	na	PF814_STE_AC_12_RN_3_12	10814	T	none	PFIT_1100300	454/573 (79%)

na: the A and B grouping was not applicable for stevor.

N/A: The sequence was not recovered in capillary sequencing data.

nd: Isolates where pir expression profiling by cloning and capillary sequencing was not done.

none: No sequence match in the cloning and capillary sequence data.

In two other patient isolates, 10735 and 10739, *rif* expression had been previously analysed by cloning and capillary sequencing (Chapter 4). For these isolates left over purified amplicon from the cloning and sequencing experiment were sequenced by 454. The quantities of the remaining amplicons was limited and therefore carrier bovine DNA was added to these samples prior to library preparation and sequencing. The resulting 454 reads and assembled contigs were contaminated with many bovine genes (identified by BLAST searches against the NCBI nucleotide nr database), as well as other P. falciparum non-pir genes and human genes. For this reason many contigs were discarded leaving few *pir* contigs for expression analysis (Table 6.1 and 6.2). Despite the limited data expression profiling was done for 2 stages in both isolates, and repertoire coverage assessed using data from the gDNA (Figure 6.11).

For the isolate 10735 expression data was obtained from rings and late ring/ early trophozoite stage (the late ring/ early trophozoite stage and early trophozoite stage was combined). As expected there was an overlap between transcripts detected by 454 amplicon sequencing and cloning and capillary sequencing data. 8 of the 9 dominant *rif* transcripts in this isolate were also detected in the capillary sequencing data. 4 *rifA* were dominantly expressed in this isolate (Figure 6.11), all of which were detected in the capillary sequence data at similar stages. One *rifA* transcript 454_RIF_1089 was dominant in rings and the other 3 were dominant in the late ring/ early trophozoite stage. For *rifB* 5 transcripts were dominantly expressed (Figure 6.11), 3 dominant in rings, 454_RIF_0518, 454_RIF_0755, and 454_RIF_0993. Of these only 454_RIF_0993 was recovered in rings in the capillary data (RIF_K017) at low levels. For the transcript 454_RIF_0518 one read was recovered in trophozoites in the capillary data, while 454_RIF_0755 was not recovered at all. 3 *rifB* were dominant in

the late ring/ early trophozoite stage, 454_RIF_0803, 454_RIF_0906 and 454_RIF_0755. 454_RIF_0906 was detected in the late ring/ early trophozoite stage in capillary data while for the other 2 no transcripts were detected in capillary data. For the isolate 10739 very few reads were recovered from the two cDNA samples, 56 reads from rings and 162 reads from trophozoites (Table 6.2). 3 *rifA* were dominantly expressed in rings, 454_RIF_0589, 454_RIF_1943, and 454_RIF_0768, of which 454_RIF_0768 was maintained as a dominant transcript in trophozoites (Figure 6.11). Four other *rifA* were dominantly expressed in trophozoites, one of which was recovered in trophozoites in the capillary data (454_RIF_0733), while 3 were not detected (454_RIF_0733, 454_RIF_0935 and 454_RIF_0951). For *rifB* 5 transcripts dominated expression in the 2 time points, 3 unique to rings and 2 unique to trophozoites. The dominant ring transcript 454_RIF_0814 was also detected in rings in the capillary data (RIF_K122). The other 2 transcripts were not detected in any time point in the capillary data. One of the 2 dominant transcripts in trophozoites 454_RIF_1663 was detected in rings in the capillary data while the other one 454_RIF_1372 was not recovered.

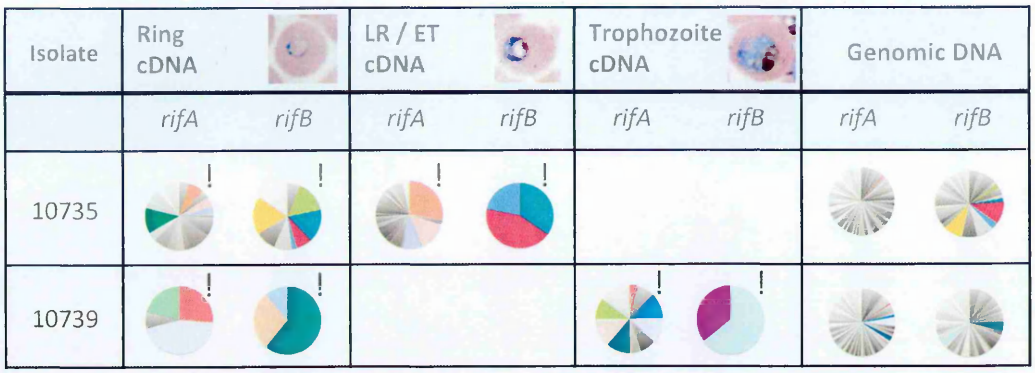


Figure 6.11. *Piecharts showing rif genomic repertoires and transcript abundance in the IDC of patient parasite isolates, 10735 and 10739 Coloured regions of the cDNA pies represent transcripts that accounted for 10% and above of the total expression in at least one time point, while in gDNA pies, coloured regions indicate genes whose read counts represent >10% of the total reads obtained from amplification of matched cDNA (see Table 6.4 below). !: Poor sequence depth (below 1000 reads).*

Table 6.4*. Dominant rif transcripts in the IDCs of 2 patient isolates 10735 and 10739*

Cluster ID	pir	Type	Cluster representative	Isolate	Stage of peak expression	Capillary sequence ID	BLAST hit	Length of overlap (% identity)
454_RIF_1089	rif	A	PF735_RIF_AC_00_MR_1	10735	R	RIF_K124	PF3D7_1000400	780/990 (78%)
454_RIF_0535	rif	A	PF735_RIF_AC_30_ET_10	10735	LR/ET	RIF_K028	PF8443_8_10_U528800	736/873 (84%)
454_RIF_0578	rif	A	PF735_RIF_AC_20_PR_1	10735	LR/ET	RIF_K019	PF3D7_0900200	792/976 (81%)
454_RIF_0620	rif	A	PF735_RIF_AC_00_MR_6	10735	LR/ET	RIF_K215	PF3D7_0833400	733/888 (82%)
454_RIF_0518	rif	B	PF735_RIF_AC_00_MR_3	10735	R	RIF_K051	PF8443_8_10_U526900	874/887 (98%)
454_RIF_0993	rif	B	PF735_RIF_AC_00_MR_13	10735	R	RIF_K017	PF3D7_1149800	655/775 (84%)
454_RIF_0755	rif	B	PF735_RIF_AC_00_GE_2_49	10735	LR/ET	none	PF8443_8_3_U486000	783/835 (93%)
454_RIF_0803	rif	B	PF735_RIF_AC_00_MR_11	10735	LR/ET	RIF_K051	PF3D7_0201000	663/778 (85%)
454_RIF_0906	rif	B	PF735_RIF_AC_00_GE_2_23	10735	LR/ET	RIF_K212	PF3D7_1041100	712/887 (80%)
454_RIF_0589	rif	A	PF739_RIF_AC_00_GE_2_11	10739	R	RIF_K407	PF3D7_1200500	739/969 (76%)
454_RIF_0768	rif	A	PF739_RIF_AC_00_MR_1	10739	R	RIF_K057	PF3D7_0200200	807/812 (99%)
454_RIF_1943	rif	AX	PF739_RIF_AC_10_RN_4	10739	R	none	PRCDC_0059300	109/118 (92%)
454_RIF_0733	rif	A	PF739_RIF_AC_40_TR_2	10739	T	none	PFSAO75_258400	704/882 (79%)
454_RIF_0935	rif	A	PF739_RIF_AC_40_TR_4	10739	T	none	PF8443_8_1_U554500	728/851 (85%)
454_RIF_0951	rif	A	PF739_RIF_AC_00_GE_2_71	10739	T	none	PF3D7_0300200	604/664 (90%)
454_RIF_1515	rif	A	PF739_RIF_AC_40_TR_9	10739	T	none	PF8443_8_12_U347300	503/587 (85%)
454_RIF_0814	rif	B	PF739_RIF_AC_00_GE_2_111	10739	R	RIF_K122	PF3D7_0401000	640/738 (86%)
454_RIF_0991	rif	B	PF739_RIF_AC_00_GE_2_88	10739	R	RIF_K272	PF8443_8_7_U438500	740/778 (95%)
454_RIF_1581	rif	BX	PF739_RIF_AC_00_MR_6	10739	R	none	PF8443_8_3_U454900	462/523 (88%)
454_RIF_1371	rif	B	PF739_RIF_AC_00_GE_2_102	10739	T	RIF_K145	PFDG_00091	644/747 (86%)
454_RIF_1668	rif	B	PF739_RIF_AC_50_MT_7	10739	T	RIF_K073	PF3D7_0900300	382/383 (99%)

6.5. Discussion.

Although efforts towards sequencing as many field isolates as possible from different geographical regions continue to be made, there still is little whole genome sequence (WGS) data from patient parasite isolates. Analysis of transcriptome sequence data generated by next generation sequencing involves mapping of reads to a reference genome. For the large variant multigene families of P. falciparum *var*, *rif* and *stevor*, mapping of reads to non-homologous genomes is complicated by the high variability in these genes resulting to poor mapping. As discussed before, sharing of *rif* and stevor genes across parasite isolates is limited, especially if they are from different geographical locations or between laboratory and field isolates (Chapter 3). It is therefore difficult to analyse next generation *pir* amplicon data by mapping to non-homologous genomes. In this section, an alternative approach to analyse the data was adopted, where 454 amplicon sequence reads were assembled into contigs using the assembly tool Newbler. The assembled contigs were then clustered using an online tool CD-HIT to identify unique variants. Using this approach it was possible to estimate repertoire coverage from the number of unique contigs recovered per isolate, and using the number of reads that went into generating each contig it was possible to estimate sequence depth, and therefore expression for each contig.

A major limitation of this analysis was the recovery of short sequence reads from suboptimal 454 sequence runs (discussed in Chapter 5). This resulted in assembly of short contigs that did not span the length of *rif* or *stevor* amplicon (typically about 900 and 600 nucleotides respectively). Because of the short length of the contigs it is possible

that contigs that span non-overlapping regions of the same *pir* gene might have been classified as unique, while in actual sense the contigs represent the same gene. To address this problem, further clustering of contigs was done using BLAST searches, where contigs that had high identity match to the same gene from BLAST searches the contigs were clustered into one group. The limitation with this approach was the isolate-specific *pir* variants that did not result in any significant match in the BLAST search. Another limitation in the analysis though minor is linked to the CD-HIT clustering strategy in order to identify unique variants and estimate *pir* repertoires. This approach assumes that there are no *pir* paralogs within an isolate when actually there are a few paralogs, as seen in the 3D7 genome. In 3D7 three sets of *rif* paralogs exists; PF3D7_0631700, PF3D7_0701300 and PF3D7_1101400 at 99.9% pairwise identity, PF3D7_0901100 and PF3D7_1301000 at 99.5% pairwise identity and PF3D7_0222400 and PF3D7_1040000 at 98.4% pairwise identity. It is therefore possible to underestimate the number of unique *pir* in an isolate using this clustering approach.

Another significant problem with the data generated from this analysis was errors from the 454 sequencing as well as errors in assembly, which resulted in many reads without clear open reading frames. It was difficult to tell apart genuine pseudogenes from contigs with sequence and/or assembly errors. In addition, because of the short contigs with missing segments it was not possible to classify a small number of contigs *rif* genes into the A or B subgroups. For contigs that could not be classified because of missing sequence regions or because of sequencing errors, the A and B grouping was assigned based on the BLAST top hit's group. The *rif* and *stevor* universal primers were highly

degenerate and amplified non-target genes including other *P. falciparum* genes like ribosomal RNA and proteins of unknown function. Reads from these non-target genes resulted into non-target contigs that required to be filtered out of the data, reducing read coverage for target genes.

Overall *pir* genomic repertoire data and expression data for most time points for all the patient isolates was available. From the data isolate-specific *pir* expression patterns were observed. Different *rif* genes dominated expression at the different stages with few overlapping time points. This observation was different from the expression patterns in laboratory and culture-adapted wild isolates from our previous analysis, where the dominant *rif* transcripts were expressed over multiple stages. *Stevor* expression was stable across time points for most isolates as observed in laboratory and culture-adapted wild isolate.

For isolates where *pirs* expression was analysed by both capillary and 454 sequencing differences were observed in the results detected in this analysis were not recovered in the amplicon cloning and capillary sequencing data (Chapter 4) from the same isolates. This was due to the difference in sequencing depth between the two approaches. Much more coverage was observed with the 454 sequencing and therefore more *pir* variants were detected. For *rif* the cloning bias detected in the capillary data contributed to the difference in results obtained by the two approaches. The capillary data was biased to recovery of short, mostly *rifB* sequences. This bias was not experienced using the clone-free 454 amplicon sequencing approach. Overall all the sequences recovered in the

cloning and capillary sequencing experiment were detected in the 454 sequencing data with many of the dominant *rif* transcripts in the 454 sequencing data observed in similar stages in the capillary data. For *stevor* the use of different primers between the 454 and capillary sequencing introduced differences in the results. The primers used in capillary amplicon sequencing had significant bias (Chapter 4), resulting in recovery of few stevor variants. Improved primer coverage and sequence depth in the 454 data resulted in recovery of all stevor genes belonging to the major group, *stevorA* group. The minor group is more divergent (Chapter 3) and none of the *stevor* genes were captured by our degenerate primers.

The recovery of a conserved pseudogene predominantly in cDNA in all the patient isolates analysed here was unusual. This was reminiscent of the stevor capillary sequencing results where a different *stevor* pseudogene dominated expression in all three patient isolates (Chapter 4 figure 4.9). These two stevor were different with 72.4% pairwise identity and having different orthologs in 3D7 (PF3D7_0401500 for the 454 stevor pseudogene and PF3D7_0102100 for the capillary sequence data). The expression of these conserved pseudogenes might suggest some importance of pseudogenic transcripts for example in regulation of gene expression. That the pseudogenes are highly conserved across isolates might suggest that these genes are not under selection pressure.

The non-mapping analysis approach in this chapter enabled exploration of *rif* and *stevor* expression in patient isolates without WGS data. The ex vivo parasite isolates provide a means to explore *pir* gene expression patterns *in vivo*, and what genes might be involved

in host-parasite interactions during malaria infections. The identified major *pir* transcripts may be involved in immune evasion or cytoadhesion of the IE and can be targeted for functional studies in future. Although a major limitation was the huge number of errors in the 454 sequences resulting in contigs with inframe stops, and additional step for error corrections will be necessary to improve this analysis approach.

CHAPTER 7

7. Expression profiling of a specific *rif* gene PFL2585c.

7.1. Introduction

Typically VSAs show low sequence homology across different isolates. Few exceptions exist such as the conserved *var* gene *var2csa* (Templeton, 2009a). Although the *rif* family is hypervariable, strain-transcendent variants have been identified in this study (Chapter 3) and in previous studies (Wang et al., 2009, Joannin et al., 2011, Claessens et al., 2011). One particular PFL2585c (PF3D7_1253700) is highly conserved in laboratory and wild isolates, with 99% pairwise identity between 3D7 and HB3, and 96% pairwise identity between 3D7 and IT (Claessens et al., 2011). A previous study looking at ex vivo transcriptomes of patient isolates showed PFL2585c to be significantly up regulated in relation to the other *rif* genes in the patient isolates by microarray and real time PCR (Lemieux et al., 2009).

Analysis of *rif* expression by RT-PCR using the complex *rif* primers revealed clear temporal pattern of expression across the IDC (Chapter 5 and 6). For certain genes, peak expression was in the early stages of the IDC, for some peak expression was the middle or late stages. The expression patterns could give an indication as to the role played by specific variants, for example transcripts that come on in late stages might play a role in invasion while transcripts that have maximal expression in early to mid-IDC could play a role in immune evasion of the IE. These results however were not able to clearly define expression patterns for genes expressed at low levels or genes that are not well captured

by the degenerate complex *rif* primers, such as PFL2585c. Gene-specific PCR would better elucidate expression profiles for such genes. For this reason reverse transcriptase PCR (RT-PCR) was employed to profile expression for this specific *rif,* PFL2585c, across IDCs in laboratory and wild isolates in order to identify peak expression time points. This *rif* is the focus of an on going study within our research group aimed at determining whether RIFIN HVLs are exposed to host antibodies on the infected red cell surface (Kibwana E., et al, manuscript in preparation). PFL2585c hypervariable region from 3D7 gDNA was expressed as a recombinant protein with a FLAG tag. This recombinant antigen was used to purify antibodies from highly reactive human sera from immune Kenyan adults. The aims of the experiments in this chapter were to identify Kilifi isolates that expressed PFL2585c, and identify the peak expression point in order to target these isolates for surface localisation studies in live immunofluorescense assays using the anti-PLF2585c antibodies.

7.2. Objectives

i.) To identify PFL2585c orthologs in wild isolates sampled from patients in Kilifi.

ii.) To identify wild and laboratory isolates expressing PFL2585c and identify peak expression time points in these isolates.

iii.) To determine whether RIFIN HVRs are accessible to antibodies on the surface of IE.

7.3. Methods

7.3.1. Retrieval of PFL25865c orthologs from laboratory and wild isolates.

PFL2585c sequences were retrieved from PlasmoDB (3D7), whole genome sequence data from the WTSI (HB3, IT and 3 Kilifi isolates), from published data for the Brazilian field isolate (Albrecht et al., 2006) and from whole genome sequence data from culture-adapted Kilifi isolates (9215, 9605 and 10668) and from de novo assembled 454 amplicon sequencing data for ex vivo Kilifi isolates (8383, 10727, 10535 and 10814) as described in Chapter 6. 3D7 PLF2585c (PF3D7_1253700) was used to retrieve orthologs from other genomes by BLAST analysis against all *rif* database on Geneious® v6.1.5. Multiple sequence alignments were carried out in Geneious® using MUSCLE with default parameters.

7.3.2. Amplification of PFL25865c by reverse transcriptase (RT) PCR.

Samples used in this study include laboratory isolates 3D7, IT, HB3, A4, P4 and R29, and Kilifi patient isolates. RNA was extracted using Trizol® as described in the methods chapter (2.5.1). Reverse transcriptase PCR (RT-PCR) for RNA and regular PCR for gDNA to amplify the hypervariable region of PFL2585c was done using specific primers, PFL2585cloop-L 5' TCC CCC GGG TTA TCC CAA AGA CCA TGT ATT TGC CTC G 3' and PFL2585cloop –R 5' GGC ATG CCA TGG ACT ACA AGG ACG ACG ATG ACA AGT TAT TAA CAG ATG CTG CTG CAA AAC 3'. The PCR for PFL2585c detection was set up as follows; 2mM $MgCl_2$, 0.4uM dNTPs, 0.4uM PFL2585c loop primers, 1X colourless GoTaq® PCR reaction buffer (with final $MgCl_2$ concentration of

1.5mM) and 1 unit of GoTaq® polymerase. 2ul of cDNA or 1ul of gDNA was used in a final reaction volume of 50ul. PFL2586c amplicons for sequencing were generated using Platinum® Taq HiFi the PCR was set up as follows: 2mM MgSO4, 0.4uM dNTPs, 0.4uM PFL2585c loop primers, 1X HiFi buffer, and 0.5 units of Platinum® Taq HiFi. PCR was performed for 35 cycles under the following conditions: 94^0C for 30 seconds, 64^0C for 30 seconds and 68^0C for 60 seconds.

PFL2585c amplicons from Platinum Taq PCR were purified prior to cloning and sequencing as described in the methods chapter (2.6.4 and 2.6.5). Sequence analysis was done in BioEdit Sequence Alignment editor (v 7.0.5.3) and Geneious Pro (5.3.4-6.1.5) as explained in the methods chapter (2.9.1).

PFL2585c HVR amplicons from 3D7 gDNA were cloned into an expression vector pEU-E01 for RIFIN protein expression using the cell free wheat germ expression system (Tsuboi et al., 2008). The cell-free expression system overcomes many limitations of using bacterial cell expression systems for *Plasmodium* proteins including insolubility, toxicity and incorrect conformation of expressed proteins.

Table 7.1. PFL2585c sequences from laboratory and wild isolates.

Isolate	sequence ID	Sequence name	Length	Source
3D7	3D7	PF3D7_1253700	1,077	PlasmoDB[a]
IT	IT	PFIT_0801500	1,077	WGS
HB3	HB3	PfHB3_BroadWash515900	1,071	WGS
Brazilian isolate	F1	DQ265365.1	991	Published data[b]
9215	K1-G	PF8443_8_4_U509500	1,071	WGS
9605	K2-G	PF8443_8_6_U541700	1,071	WGS
10668	K3-G	PF8443_8_8_U487800	1,071	WGS
8383	K4-A	PF383_RIF_AC_00_GE_2_33_PFL	889	454 amplicon sequencing
10727	K5-A	PF727_RIF_AC_00_G1_2_90_PFL	551	454 amplicon sequencing
10735	K6-A	PF735_RIF_AC_00_GE_2_88_PFL	745	454 amplicon sequencing
10814	K7-A	PF814_RIF_AC_00_GE_2_36_PFL	858	454 amplicon sequencing
10761	K8-A	PF761_RIF_AC_00_ER_2	895	454 amplicon sequencing

a: PlasmoDB 13.0 release.

b: (Albrecht et al., 2006)

7.3.3. Live immunofluorescence assays (IFA).

As previously described anti PFL2585c antibodies were purified from immune sera using PFL2585c HVRs expressed as FLAG-tagged protein as previously described (Chapter 2.12.1) (Kibwana E., Kadasia K. et al., manuscript in preparation).

Early to mid trophozoites from parasites identified as high expressers by RT-PCRs were used in the IFA as previously described (Chapter 2.12.2). Briefly 500ul of the cultures were spun down and washed three times in incomplete media prior to incubation with the antibody solution at 37^0C for 30 minutes to an hour in a 1:1 ratio of iRBC and purified antibody solution in 1% BSA solution. The pellets were then washed three times in 1000ul cold PBS with 1% BSA and then incubated with 50ul of the secondary antibody, goat anti-human IgG Alexa 488 F(ab)' conjugate at 37^0C for 30 minutes in the dark. In the last 5 to 10 minutes of incubation with secondary antibody the nuclei stain Hoechst/ DAPI at 1μg/ml was added, 50ul of a 1:1000 DAPI/Hoechst (1μg/ml, diluted in 1% BSA/PBS). The pellets were then washed three times in1000μl of cold PBS with 1% BSA, resuspended in 50μl of 1% BSA in PBS and 20μl of the washed pellet placed on a glass slide and sealed with a coverslip. The slides were examined using 40x or 100x (oil) objective, and incident light of 450-49nm for FITC fluorescence, 390-440nm for DAPI/Hoechst.

7.4. Results

7.4.1. High sequence conservation of PFL2585c in laboratory and wild isolates.

In addition to 3D7 10 orthologs from IT, HB3, a Brazilian isolate and 7 Kilifi isolates. Multiple sequence alignment on the 11 sequences showed high sequence homology even within the Hypervariable region (Figure 7.1). The overall pairwise identity of PFL2585c orthologs is 98.2% over the full length of the sequences. The actual mean pairwise identity is probably higher as most of the disagreement was at the sequence ends of three wild isolates probably due to PCR and sequencing errors. Sequences from Kilifi isolates were obtained by cloning and capillary sequencing PFL2585c amplicons from cDNA or gDNA (an example of gel purification of amplicons from 10594 in Appendix Figure 9.4). The mean pairwise identity within the HVL in laboratory and wild isolates was 99.2 % (Appendix Figure 9.5).

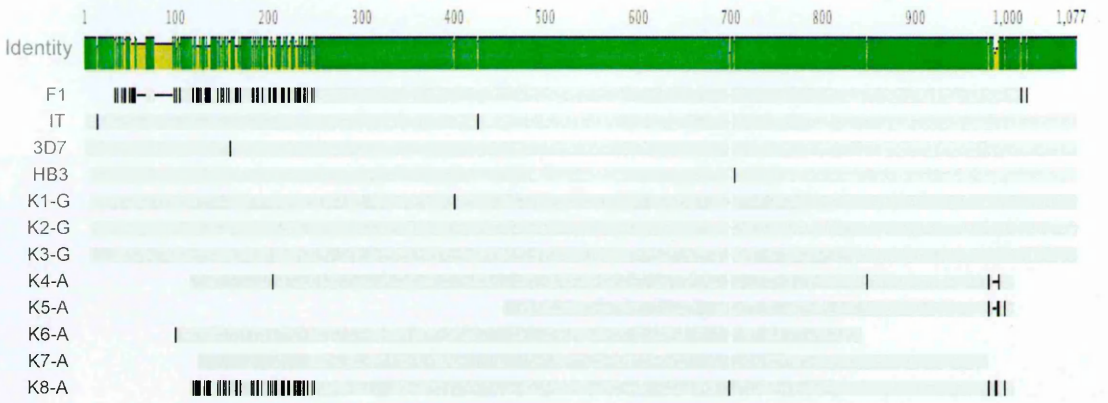


Figure 7.1. *Multiple sequence alignments of PFL2585c and 10 orthologs from IT, HB3, a Brazilian isolate (F1), and 7 Kilifi isolates, 9215 (K1-G), 9605 (K2-G), 10668 (K3-G), 8383 (K4-A), 10727 (K5-A), 10735 (K6-A) and 10814 (K7-A). The mean pairwise identity over all the pairs in the column is shown above the sequence alignment. Green: 100% identity. Green-brown: at least 30% but below 100%. Black vertical lines indicate disagreement between sequence in alignment and the consensus sequence.*

7.4.2. PFL2585c expression profiling by RT-PCR in laboratory isolates.

To profile expression of a specific *rif* PFL2585c specific RT-PCR was used to detect transcripts in laboratory and wild isolates. Specific primers were used to amplify PFL2585c from cDNA from 6 timepoints in 3D7, HB3 and IT and from ring and trophozoite cDNA from A4 laboratory lines. PCR on gDNA from the same samples was included as a positive control for PCR. PFL2585c transcripts at the expected size, 400bp, were detected in HB3, IT and A4 (Figure 7.2). Different PFL2585c expression profiles were observed in different laboratory isolates. In HB3 transcripts were detected in most stages of the IDC, peaking in late rings and going off in schizonts (Figure 7.2. A II). In IT PFL2585c expression was restricted to small window between late rings and early trophozoites, staying very low in later stages (Figure 7.2. B). In A4 PFL2585c transcripts were mainly detected in schizonts (Figure 7.2. C) but low intesity band could be seen in trophozoites. PFL2585c was not expressed in 3D7 as no signal was detected in 3D7 cDNA (Figure 7.2. A. I and Appendix Figure 9.3). Expression in HB3 was validated by results from total *rif* amplicon deep sequencing. Reads mapping to the PFL2585c ortholog in HB3, PfHB3_BroadWash515900 were recovered in amplicons from late rings to schizonts in the 454 data (Appendix table 9.8), with most reads mapping from amplicons from schizont stage. Very few reads from 3D7 and IT cDNA mapped to PFL2585c and it's ortholog in IT, PFIT_0801500, an average read count 1.3 in 3D7 and 1.6 in IT across time points (Appendix table 9.8).

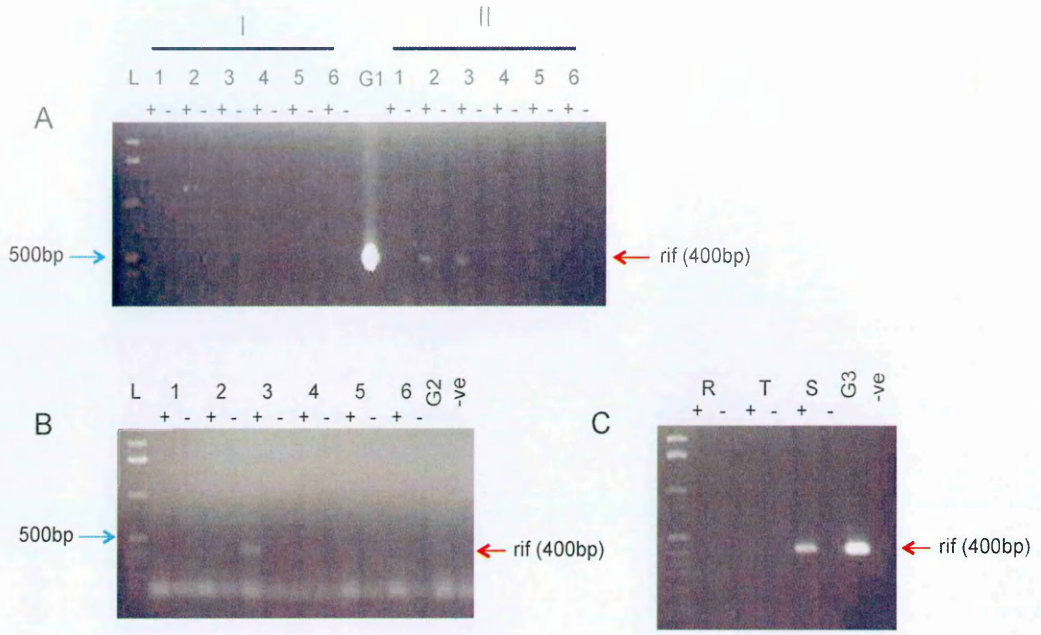


Figure 7.2. *1.5% agarose gel electrophoresis images of PFL2585c amplicons in laboratory isolates. (A) Amplicons from 3D7 cDNA at 6 time points (I) and 3D7 gDNA (G1), and HB3 at 6 cDNA time points (II). (B) PFL2585c amplicons from IT cDNA at 6 timepoints and fro IT gDNA (G2). (C) PFl2585c amplicons from A4 ring (R), trophozoite (T) and schizont (S) cDNA and A4 gDNA (G3). L:1 kb plus DNA ladder. + and – indicate RT enzyme positive or negative cDNA samples respectively.*

7.4.3. PFL2585c expression profiling by RT-PCR in Kilfi isolates

In previous literature it was observed that PFL2585c expression was higher in patient isolates as compared to 3D7 (Lemieux et al., 2009). Expression profiles were analysed in culture-adapted isolates as well as ex vivo patient isolates. For the cultured-adapted wild isolates, PFL2585c specific RT-PCR was carried out on timepoint RNA samples from 4 IDC stages rings, late rings to early pigmented trophozoites, mid to late trophozoites and schizonts. PFL2585c expression was also analysed in an additional high rosetting laboratory clone R29 at four timepoints same as the cultured wild isolates.

PLF2585c was expressed at low levels in one culture-adapted patient isolate 10668, and in R29. In 10668 PFL2585c transcripts were detected in all timepoints with a peak in band intesnity in rings (Figure 7.3. A). In R29 the peak expression was in the late ring to early trophozoite stage, with very little transcription in trophozoites and schizonts (Figure 7.3. A). A very intense signal was observed in gDNA from both parasites. PFL2585c transcripts were detected in rings and trophozoites of a culture-adpated Kilifi isolate 8383 (Figure 7.3. C and Appendix figure 9.2)

SA075 R+ and 9106 did not express PLF2585c because of a deletion of this gene loci as no band was detected in gDNA samples (Figure 7.3. B). To verify lack of expression rather than cDNA degradation, a confirmatory experiment was carried using cDNA freshly prepared from the same SA075 R+ RNA samples and similar results were obtained (Appendix Figure 9.1). Genomic deletion in SA075 R+ and 9106 was confirmed using BLAST against *rif* sequences from whole genome sequence data from the two isolates. PFL2585c orthologs were not identified in the two genomes. For two other Kilifi isolates, 6816 and 8948, no signal was detceted from cDNA but it was not clear whether

the gene was switched off or deleted because no gDNA sample was available to confrim deletion (Appendix figure 9.2).

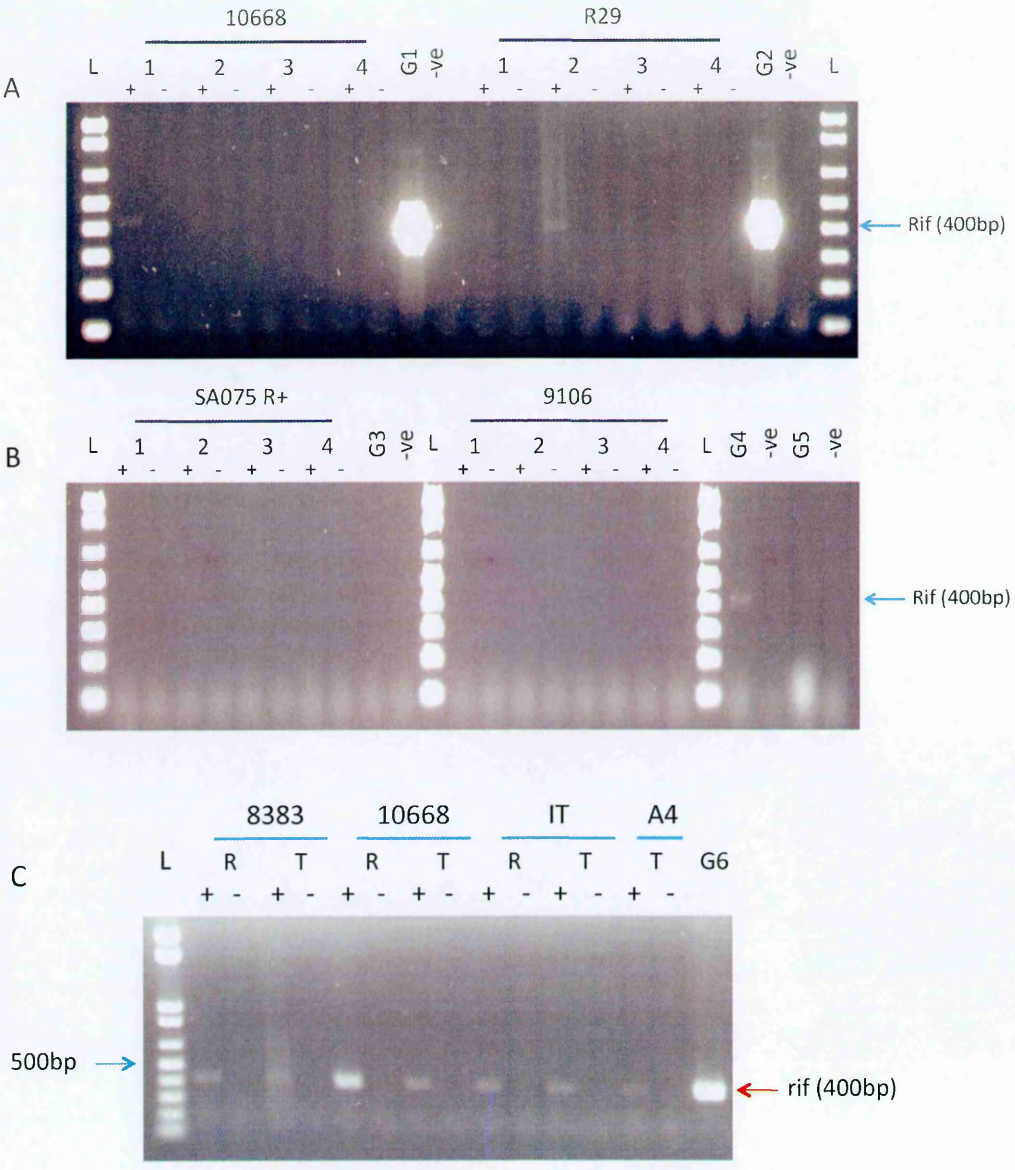


Figure 7.3. *5% agarose gel electrophoresis images of PFL2585c expression profiles in wild and cultured isolates. PFL2585c transcripts at 4 stages of the IDC, rings (1), late ring to early trophozoites (2), trophozoites (3) and schizonts (4) in culture-adapted wild isolates: (A) 10668 and R29, (B) SA075 R+ and 9106. (C) PFL2585c transcripts from rings (R) and trophozoites (T) in culture-adapted Kilifi isolates, IT and from A4 trophozoites. PCR products from gDNA are labelled as G1 for 10668, G2 for R29, G3 for SA075, G4 for 8383, G5 for 9106 and G6 for A4. L:1 kb plus DNA ladder. + and – indicate RT enzyme positive or negative cDNA samples respectively.*

Analysis of PFL2585c expression in ex vivo parasite isolates sampled from children in Kilif with non-severe malaria was assessed to determine whether patient isolates showed a much higher level of expression. For two isolates, 10594 and 10595 cDNA had been prepared directly from total RNA without prior gDNA digestion. No RNA samples were available for these isolates to repeat the experiment with DNA-free cDNA. For two other patient isolates 10735 and 10739 RNA was DNase digested prior to cDNA preperation.

The expected PFL2585c amplicon at 400bp was observed in one of the isolates, 10594 at all time points, with peak in band intensity at the fourth time point corresponding to trophozoites (Figure 7.5 A. II). A non-specific amplification product 750bp in size was detected in this isolate (Figure 7.5. A). The non-specific 750bp band was identified as human hemoglobin beta gene by sequencing the band and performing a BLAST search againts NCBI nucleotide nr database. This contaminant originated from reticulocytes in the patient sample.

No bands were detected in 10595 cDNA. PFL2585c PCR on gDNA revealed a deletion of this gene in 10595 (Figure 7.5. B). This explained why no transcripts could be detected in this isolate. Genomic DNA quality was confirmed by total *rif* PCR on all the gDNA samples. A smear of bands at 800 to 950bp corresponding to *rif* amplicons were detected in all the 4 isolates tested, 3D7, IT, 10594 and 10595 (Figure 7.5. C). This deletion in 10595 was confirmed by comparative genomic hybridisation (CGH) data from colleagues (Simam, J unpublished) and (Appendix figure 9.6).

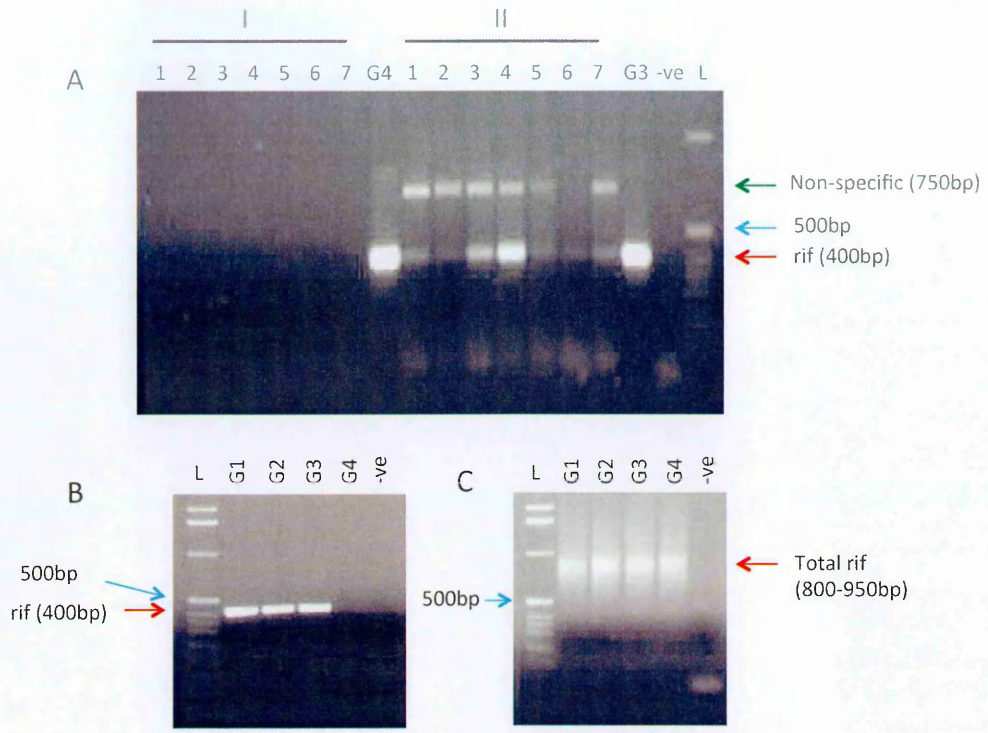


Figure 7.4. *PLF2585c expression profiling in patient isolates. (A) RT-PCR products from 10595 (I) and 10594 (II) at 7 time points across a single IDC ex vivo, and regular PCR products from 3D7 gDNA (G4) and IT gDNA (G3). (B) PFL2585c PCR products and (C) total rif PCR product from gDNA from 3D7 (G1), IT (G2), 10594 (G3) and 10595 (G4). The specific PFL2585c amplicon was detected at 400bp (red arrow) in all time points in 10594 and in all gDNA samples with the exception of 10595. The blue arrow points to the 500bp band on the DNA ladder. A non-specific amplification product was detected at 750bp in 10594 cDNA (green arrow). The expected total rif PCR product complex rif was observed at about 800-950bp in all gDNA samples. L: 1kb plus DNA ladder.*

PFL2585c expression was analysed in two other patient isolates 10735 and 10739. No transcripts were detected in 10739 (Figure 7.6. A). In 10735 very low intesity bands were detected at time point 3 and 4 corresponding to the late rings and early trophozoite stages (Figure 7.6. B). The gDNA samples were not analysed by PFL2585c-specific PCR for these two isolates, but were analysed by 454 total *rif* amplicon sequencing data. From the 454 data PFL2585c sequences were not recovered in 10739 cDNA or gDNA, implying that there was a deletion of this locus in this isolate. In 10735 PFL2585c sequences were detected in amplicons from gDNA but not from cDNA. Microarray data from our collaborators (Mackinnon) indicated that PFL2585c was expressed in 10735 with peak expression in the 3rd timepoint corresponding to late rings and early trophozoites (Appendix figure 9.7).

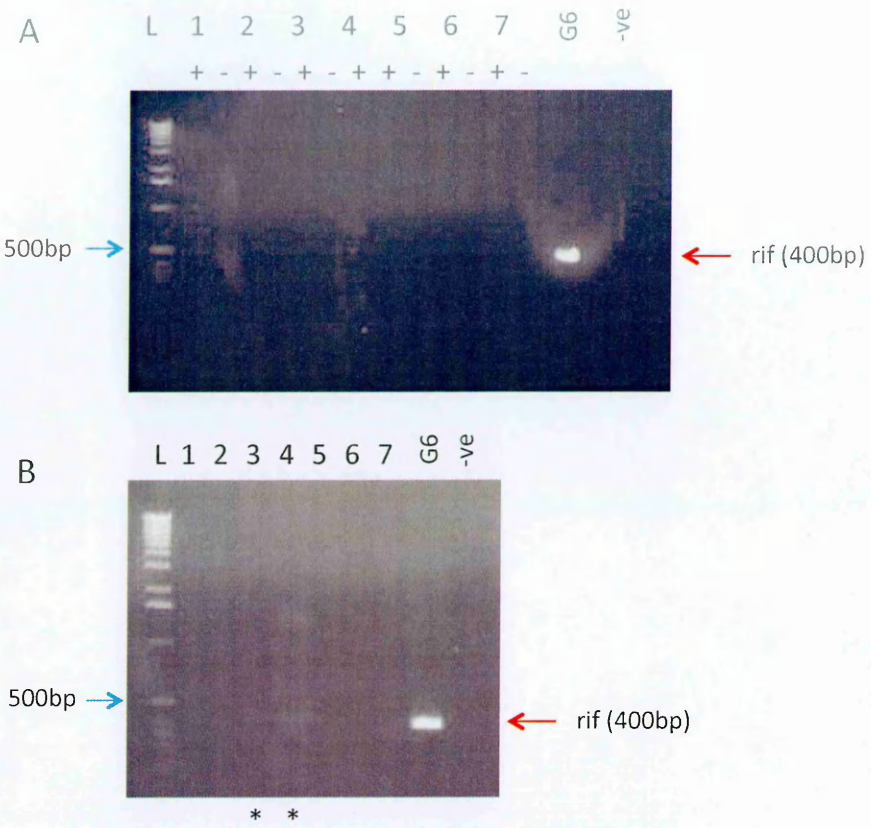


Figure 7.5. *PLF2585c expression profiling in two patient isolates 10739 (A) and 10735 (B) at 7 time points across one IDC ex vivo, and regular PCR products from A4 gDNA (G6). The specific PFL2585c amplicon was detected at 400bp (red arrow) in the gDNA samples and at very low intensity in the 3rd and 4th time points in 10735 (*). L: 1kb plus DNA ladder.*

Consolidating all the data on PFL2585c expression, PFL2585c RT-PCR, total *rif* 454 amplicon sequencing and microarray, 12 isolates out of a total of 22 were identified as expressing this specific *rif* gene. Of these 5 were laboratory lines, 3 were culture-adapted wild isolates and 4 were ex vivo patient isolates (Table 7.2). For most isolates peak PLF2585c expression was observed in late rings to early trophozoites. Only one isolate 10668 showed maximal expression in rings. For 5 Kilifi isolates, a deletion in the PFL2585c loci was detected (Table 7.2). This was supported by whole genome sequence data (for SA075 and 9106) and microarray CGH data (for 10595). Previous studies (Kidgell et al., 2006, Ribacke et al., 2007) as well as an ongoing study by our collaborators (Simam J., and Mackinnon M.) have shown that the region within which this gene lies is within a CNV and the deletion is not uncommon in Kilifi patient isolates. For 2 isolates 6816 and 8948 PFL2585c deletion could not be confirmed because we lacked parasite gDNA to test this by PCR.

Table 7.2. Summary of PFL2585c expression in laboratory and wild isolates.

PFL2585c expression summary			
Lab			
Parasite isolate	**PFL2585c Expression**	**Peak expression point**	**PLF2585c surface expression***
3D7	OFF (g) (s)	na	N/A
IT	ON (g)	Late ring/early trophozoite	YES#
HB3	ON (g)	Late ring/early trophozoite	Not tested
A4	ON (g)	Late trophozoite to schizont	Not tested
P4	ON (g)	Rings and trophozoites	Not tested
R29	ON (g)	Rings to early trophozoite	NO
Culture-adapted pateint isolates			
Parasite isolate	**PFL2585c Expression**	**Peak expression point**	
8383	ON (g)	Trophozoite	YES
10668	ON (g)	Rings	YES
9106	Deletion (g)	na	N/A
SA075c	Deletion (g)	na	N/A
9215	ON	Rings	YES
6816	OFF (b)	na	N/A
8948	OFF (b)	na	N/A
Patient isolates			
10739	Deletion (s)	na	N/A
10735	ON (?) (m)	Late ring/ early trophozoite	Not tested
10594	ON (g)	Late trophozoite	Not tested
10595	Deletion (g)	na	N/A
10593	ON (g)		
10727	OFF (s) (m)	na	N/A
10747	Deletion (s)	na	N/A
10761	ON (s) (m)	Schizonts	N/A
10814	OFF (s) (m)		N/A

a: cDNA potentially had gDNA contamination

b: no gDNA sample tested

#: Observed some non-specific staining with secondary antibody

?: Very faint band by TR-PCR, no reads from cDNA in 454 data, but on in microarray data

g : Expression data from PCR and gel

s: Expression data from total rif 454 and capillary sequence data

m: Expression data from microarray data (Rono M., and Mackinnon, M unpublished)

**: Work done by other members of our research group (Kibwana E., et al, manuscript in preparation).*

7.5. Surface labelling of PFL2585c antigens on the surface of infected erythrocytes.

The previous analysis on expression of PLF2585c by RT-PCR enabled identification of parasite isolates that showed high levels of transcripts at one or more IDC stages. The high expressers included HB3, IT, and culture-adapted field isolates 8383, 9215 and 10668. HB3 parasites in our laboratory were obtained already infected with antibiotic resistant mycoplasma and could not be cultured for IFA. IT, 8383 and 10668 parasites were cultured to above 5% parasitemia and harvested at early to mid trophozoite stages for staining. Clear surface labeling of PFL2585c with immune serum and the affinity-purified (AP) antibodies was observed in the culture-adapted Kilifi patient isolates 8383, 9215 and 10668 (Figure 7.6 and 7.7). 9215 and 10668 were sampled from patients with severe malaria, 9215 being a rosetting isolate. 8383 was sampled from a patient with non-severe malaria. Low level background staining could be seen in 8383 with the hyperimmune serum 17H (Figure 7.6 upper panel) and in 9215 with the AP anti-RIFIN antibodies (Figure 7.7. In one patient isolate, 10668, staining was observed in a rosette (Figure 7.6 lower panel). Although staining was observed in the laboratory isolate IT there was significant background staining observed on non-infected erythrocytes. It was therefore unclear whether the labeling on the IE was from the anti-RIFIN antibodies or as a result of non-specific staining. Despite a large population of cells staining with DAPI, only a small proportion was stained with the anti-RIFIN antibodies (Figure 7.7). This is an indication of the specificity of the antibodies. As observed in the RT-PCR results PFL2585c transcript signal was low for most isolates and therefore few IE were expected to express this antigen on their cell surface. No staining was detected with nonimmune

serum from UK donors, or after staining with the secondary antibody alone. Staining with immune sera from African donors resulted in many more labeled IE as compared to the number of IE labeled when using the purified anti-RIF antibodies, suggesting specificity of these antibodies.

Although surface labeling was observed for few IE in live IFA this result shows that RIFINs are localised to the surface of IE and are targets of naturally acquired antibodies giving evidence for a role in host parasite interactions including immune evasion.

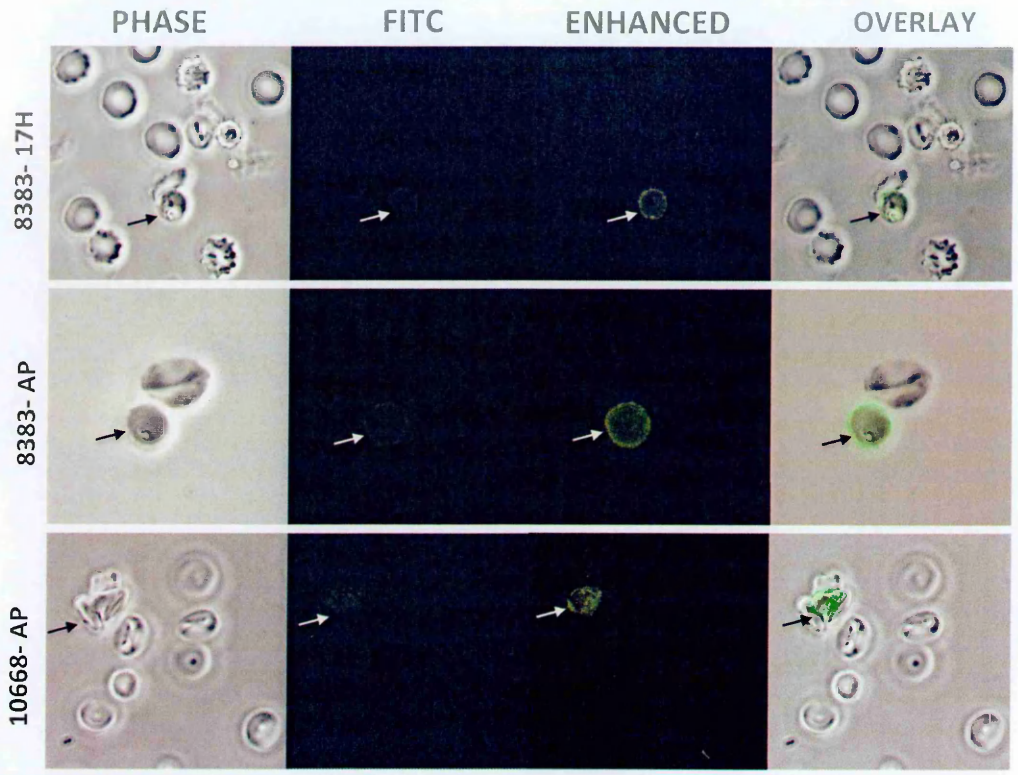


***Figure 7.6**. Live immunofluorescence analysis in trophozoite stages of two culture-adapted patient parasite isolates 8383 (upper and middle panel) and 10668 (lower panel), using a hyper immune serum 17H (upper panel) and affinity purified anti-PFL2585c antibodies (middle and lower panels).*

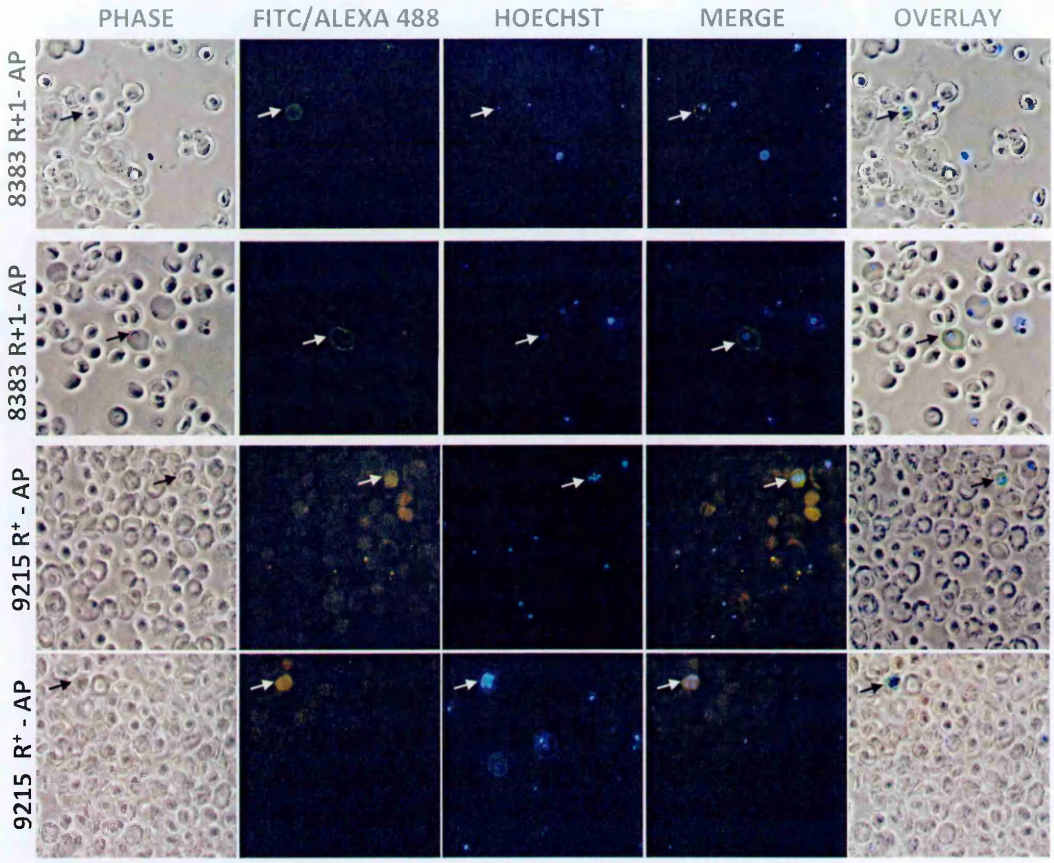


Figure 7.7. *Live immunofluorescence analysis in trophozoite stages of two culture-adapted patient parasite isolates 8383 (upper panels) and 9215 (lower panels), using the affinity purified anti-PFL2585c antibodies. The nuclei were stained with DAPI (blue).*

7.6. Discussion.

PFL2585c orthologs were identified in all the laboratory lines analysed and in 8 Kilifi isolates with high sequence conservation even within the HVL (99.3%).

It was clear that PFL2585c was not the predominant *rif* expressed in the isolates sampled as evidenced by low intensity bands from RT-PCR, in comparison to total *rif* bands (Appendix Figure 9.3). Low level expression could be detected with the highly sensitive RT-PCR in 12 out of 22 isolates. A larger proportion of laboratory isolates expressed this gene as compared to pateint isolates. PFL2585c lies within a CNV region, therefore deletions of this gene is expected. There was a high frequency of deletions of this region in ex vivo isolates (5 out of 16 isolates with a deletion) but not in any of the laboratory lines. This deletions could explain the low frequency of PFL2585c expressers from patient isolates. In the isolates that expressed PFL2585c we observed different peak expression time points in different isolates. Low level transcription in HB3 and 10668 was observed in the 454 amplicon sequencing results (14.2 and 1.5 average reads mapped from all timepoints in HB3 and 10668 repsectively), though this could also be due to poor coverage of this gene by the degenerate *rif* primers.

Other members in my research group have expressed the hypervariable domain of PFL2585c as a recombinant fusion protein containing a FLAG tag. Sera from adults and children in Kilifi reacted to this recombinant antigen. The antigen was then used to purified antibodies from immune sera from the adults. Trophozoite stage parasites from 3 isolates identified as expressing the PFL2585c gene, by RT-PCR, IT, 8383 and 10668,

were obtained for immunofluorescence assays (IFA). The purified antibodies were able to label native antigen on the surface of infected erythrocytes from IT and three Kilfif isolates in live IFA (Figure 7.6 and 7.7 and Table 7.2), giving evidence of PFL2585c role in antigenic variation of the infected erythrocyte (Kibwana E., et al, unpublished). Althoug HB3 was one of the highest PFL2585c this parasite line was not used as it was contaminated with antibiotic resistant mycoplasma.

Previous work on localisation of RIFINs showed that type A RIFINs were localised out of the parasitophorous vacuole and into the maurer's clefts and were associated with the IE surface. Type B RIFINs however remained within the parasite vacuole and did not localise to the maurer's clefts (Petter et al., 2007). This differential localisation that would suggest a divergence in function in the two RIFIN subgroups was shown for a laboratory parasite lines. It has been suggested that VSAs are more highly expressed in patient isolates than in laboratory isolates (Daily et al., 2005, Bachmann et al., 2012) and therefore it would be useful to repeat RIFIN localisation studies with patient isolates to validate that B type RIFINs are not exported to the IE surface.

Future work will involve quantifying expression in larger numbers of patient isolates using quantitative real time PCR (qPCR), to ascertain whether patient isolates have higher gene expression levels. In addition qPCR will be used to determine whether expression of this gene is associated with any clinical phenotype by using patient isolates with clearly defined clinical phenotypes. The purified antibodies were only able to label a small population of infected erythrocytes in three culture adapted patient isolates but this

provides good evidence that RIFINs are coexpressed with PfEMP1 on the surface of infected red blood cells and are targets of naturally acquired immunity to malaria. These results gives evidence for a role of RIFINs in host-parasite interactions including immune evasion and cytoadhesion.

CHAPTER 8

8. Discussion and conclusions

Variant surface antigens of P. falciparum modify the antigenic and cytoadhesive characteristics of the infected erythrocyte (IE) therefore play an important role in disease pathogenesis and acquisition of immunity to malaria. These roles have been mainly ascribed to the *var* encoded PfEMP1. Recent literature suggests that other VSAs are coexpressed on the IE surface and may complement the functions of PfEMP1. The *rif* and *stevor* multigene families encode the VSAs RIFIN and STEVOR. These genes belong to the larger *pir* multigene family found in other Plasmodium species. The objectives of this work were to explore sequence diversity and expression profiles of *pir* multigene families of *P. falciparum* in natural malaria infections, and to determine whether RIFIN proteins are targets of naturally acquired immune antibodies on the surface of infected erythrocytes. Most studies looking at sequence diversity of *rif* and *stevor* genes have been focused on laboratory isolates because of scarcity of data from patient isolates. Very few studies have analysed *pir* expression patterns in patient isolates.

Exploration of *pir* sequence diversity was carried out in Chapter 3. Phylogenetic studies have demonstrated that *rif* genes can be classified into functionally divergent groups and that the groups are conserved across isolates (Joannin et al., 2008, Petter et al., 2007). The *stevor* genes subdivide into 2 main groups but no evidence of functional divergence exists (Joannin et al., 2011, Schreiber et al., 2008). Phylogenetic approaches were used to classify *rif* and *stevor* genes from parasites geographically diverse origins. *Pir* repertoires were conserved in terms of number of sequences and subgrouping across isolates from

distinct geographical locations, including Kenyan isolates. Conserved groups as well as highly conserved orthologs were identified. *Pir* genes from Kenyan isolates grouped into previously described conserved groups. The *rif* genes grouped into the two major groups *rifA* and *rifB* as previously observed, with *rifB* falling into 3 subgroups, B1, B2 and a diverse B3 group (Joannin et al., 2008, Joannin et al., 2011). Different sequence subgrouping was observed when different regions of the genes, N terminal, and the hypervariable regions, were phylogenetically clustered. This is an indication of the different selection pressures on the different regions, the HVR under diversifying selection while the N terminal regions are constrained to limited variability. The centrally located *pir* genes form a conserved group, as they lie within stable regions of the genome with infrequent recombination events. These central *pirs* have orthologs in wild isolates from different geographical locations. Conserved *rif* and *stevor* genes have been identified in laboratory and wild isolates, some of which have ancient origins as evidence by having orthologs in *P. reichenowi*.

To characterize expression profiles in different parasite isolates, gene family-specific reverse transcriptase (RT) PCR was employed using sets of universal *rif* and *stevor* primers. The universal *pir* primers were able to amplify *rif* and *stevor* genes from gDNA and cDNA from laboratory and wild isolates. Although the primers were designed on 3D7, they showed good coverage of *pir* repertoires in IT and Kenyan patient isolates, with slightly reduced coverage in HB3. This might be due to the probable close relationship between 3D7, an African isolate and the Kenyan isolates, whereas HB3, a laboratory isolate originally from Honduras might not be as closely related.

As a first step in characterizing expression profiles of *pir* genes, *rif* and *stevor* amplicons were cloned and sequenced using capillary Sanger sequencing approach as described in Chapter 4. The results of this approached revealed temporal *pir* expression patterns that were parasite-specific. There was high *rif* transcript diversity within an isolate across time points and across isolates. There was little sharing of transcripts across isolates, with only 2 *rif* transcripts shared in 2 patient isolates. Dominant *rif* and *stevor* transcripts were observed in different timepoints. For *rif* genes a more focused expression to a single dominant variant was seen in the schizont stage. There was clear differential upregulation of *rif* and *stevor* genes in the rosetting clone and the non-rosetting fraction of a rosetting patient isolate. For *stevor*, primer bias resulted in preferential recovery of a stevor pseudogene from all the patient isolates.

In Chapter 5 and 6 a more in depth analysis of expression profiles in laboratory and patient isolates was carried out using next generation deep sequencing of amplicons on the 454 platform. Two approaches were used to analyse the 454 reads, 1) mapping of reads to homologous genomes and 2) de novo assembly of reads into contigs for isolates with no whole genome sequence data. High transcript diversity and little sharing of expressed sequences across parasite isolates as seen in the capillary data was confirmed. The temporal pattern of *pir* expression was more clearly observed in the deep sequence data. Stage-specific dominant *rif* and *stevor* transcripts were observed, for most isolates different transcripts dominated expression at rings trophozoites and for *rif* in some cases different sets came up in schizonts. A previous study of VSA expression profiling the

same dominant transcripts were observed at the two different peaks of expression, in rings and trophozoites (Bachmann et al., 2012). Our study had much greater sequencing depth and better primer coverage, which could explain the difference in results. To analyse the effect of selection pressure on *pir* gene expression, profiling was carried out in patient isolates that had undergone different selection treatments, 1) *ex vivo* parasite lines to test effect of host immune selection pressure, 2) adaptation to culture conditions to test effect of culture adaptation and 3) parasites selected for binding phenotypes, binding to endothelial cells (HBEC) and to uninfected erythrocytes (rosetting). Different *rif* and *stevor* transcription profiles were observed in isolates put under different selection pressures. There was more even expression in ex vivo parasites that have been under host immune selection pressure but in culture expression was focused to a few variants that were dominantly expressed. Different sets of transcripts were upregulated after culture-adaptations, and different sets after selection for HBEC binding or rosetting. In previous microarray analysis of VSA expression in laboratory parasites selected for rosetting or binding to HBEC lines, upregulation of *UpsArif* genes neighbouring the upregulated *UpsAvar* genes was observed (Claessens et al., 2012, Claessens et al., 2011). This link in *rif* and *var* expression was not clear in our results. The upregulated *rif* genes after selection for binding did not belong to the conserved *upsArif* group by BLAST analysis. It was not possible to determine whether the upregulated *pirs* were in close proximity to the dominantly expressed *var* genes because of missing chromosomal location data for the genomes of wild isolates.

In the Northern blotting analysis full length *rif* transcripts were detected in IT and a patient isolate from Kenya using RNA probes generated from homologous cDNA form these isolates. The low sensitivity of the non-radioactive labeled Digoxygenin probes meant that a lot more RNA was required to reach detection limit, and longer exposures of the blots during visualization, and both these would result in even higher background signal from ribosomal RNA, therefore more difficult to detect *rif* mRNA signal. Therefore it was not possible to carry out temporal expression profiling by northern blotting in this study. To overcome these limitations future analysis beyond the scope of this thesis will utilize purified mRNA rather than total RNA to enable detection of less abundant transcripts.

To explore whether transcripts detected by RT-PCR were expressed into protein that were localized to the surface of infected erythrocytes, and whether these native antigens were recognized by immune sera, the experiments described in Chapter 7 were carried out. A unique and strain-transcendent *rifA* PFL2585c was chosen because of its conservation in laboratory and wild isolates, and because previous literature had shown this gene to be upregulated in field isolates (Lemieux et al., 2009). PFL2585c was expressed in most of the laboratory isolates (5/6) and few wild isolates from Kenya (7/16) tested but was switched off in some isolates including the laboratory isolate 3D7. As this gene lies within a CNV region a few isolates had a deletion of this gene (5/17). Unique expression profiles were observed with peak expression occurring at different time points in different isolates. To test whether RIFIN variant domains are accessible to antibodies at the IE surface and whether they are targets of naturally acquired protective antibodies

isolates identified to express this *rif* by RT-PCR were cultured for localization analysis by live IFA. First the hypervariable domain of PFL2585c was generated as a FLAG-tagged recombinant protein. This antigen was used to affinity purify naturally occurring human antibodies from highly responsive sera from immune Kenyan adults. The antibodies labeled a minor population of IE from parasites that were identified to transcribe this gene at a population level in live IFA. This study contributes to the existing evidence for surface localization of A-RIFIN, and their role in immune evasion of IE.

8.1. Limitations of the study

8.1.1. Incomplete coverage of *pir* repertoires by primers.

A major limitation of this study was the inability to capture all *rif* and *stevor* variants using universal primers. Although coverage was good of most isolates (about 70% coverage) certain key genes were missed out. The divergent strain-transcendent *rif* PFD0070c which has been previously shown to be the dominant *rif* transcript in 3D7 was not captured. The *stevorB* subgroup was completely missed out by stevor primers. In addition our RT-PCR approach was only semi-quantitative. Therefore there remain gaps in expression profiling of *pirs* in terms of coverage and quantification. Future studies would benefit from using approaches such as RNAseq to obtain whole transcriptome sequences for *pir* expression profiling. To validate the dominant transcripts observed quantitative realtime PCR will be carried out in future studies.

8.1.2. Short reads and sequence error in the 454 amplicon sequencing results.

Suboptimal results obtained from the 454 amplicon sequencing resulted to recovery of short reads and this proved a major shortcoming during de novo contig assembly. In addition there were many errors in the sequences. These resulted in short assembled contigs with many sequence errors complicating analysis of variant types by our clustering approaches. Therefore it is possible that we overestimated the number of pir variants in the de novo assembly data. Future analysis beyond this thesis will involve error corrections for the assembled contigs and mapping sequence reads to pseudogenes generated from assembled contigs.

8.2. Future work

The direct functions of *pir* genes in malaria disease pathogenesis and in acquisition of natural immunity to malaria still need to be explored. A direct role of the related *cir* multigene family in virulence of the rodent malaria parasite *P. chabaudi* has been demonstrated (Spence et al., 2013). In the study mosquito transmitted parasites were less virulent than serially blood passaged parasites. This was dependent on an intact immune response, and was attributed to differences in *cir* gene expression in the two parasites. The mosquito transmitted parasites expressed many different cir variants, and in an animal with an intact immune response, the parasitemia was controlled, resulting in a low parasitemic chronic infection. On the other hand, serially passaged parasites expressed one dominant *cir* gene. These parasites induced the immune system towards a more proinflammatory response, and grew to uncontrollably high parasitemia resulting in

virulence. The mechanisms of virulence may be the same for *P. falciparum*, in that parasites with high homogeneity of *pir* expression may trigger more inflammatory responses that are unable to keep parasitemia in check, resulting in severe malaria. The significance of this findings is that vector transmission regulates parasite gene expression in asexual blood stages, which impacts on the host immune response to the parasite and therefore on parasite virulence. It is therefore the interplay between vector, parasite and host immune system that seems to underlie parasite virulence. Mosquito transmission regulates virulence of he parasite in the mammalian host, by modifying variant gene expression, where the mosquito-transmitted parasites express many variants as observed P. chabaudi in the aforementioned study (Spence et al., 2013) and in P. falciparum (Peters et al., 2007). The mosquito part of the parasite life cycle therefore appears to be crucial for regulating the inherent virulence of the parasite in the mammalian host, by acting like a reset switch ensuring that every new infection is not transmitting highly virulent parasites from one person to the next. These parasites with attenuated virulence can be cleared by the immune response of an immune host, but in a non-immune host the parasitemia increase uncontrollably resulting in severe disease. In *P. falciparum*, virulence has been mainly associated with *var* genes. Expression of certain groups of *var* genes has been associated with severe malaria (Warimwe et al., 2009, Salanti et al., 2003). This link is missing for *P. falciparum pir* genes. Transcription profiling of *pir* genes in *P. falciparum* isolates sampled from patients with different malaria clinical phenotypes will enable identification of *pir* genes associated with severe disease or low host immunity.

Attempts to understand the function of rif and stevor genes relies on the ability to properly classify them into distinct groups for which expression profiles can be associated with specific phenotypes such as binding to endothelial cells or rosetting. Phylogenetic trees are not appropriate for clustering highly variable and recombining sequences such as rif and stevor. This is supported by the limitations observed in chapter 3, where the resultant phylogenetic trees had low statistical support, low branch support and long branch lengths. In addition we observed phylogenetic incongruence where different regions of the genes, N-terminal, HVR and full length sequences resulted in different clade structures. With these limitations, phylogenetic trees can only be used as an initial step in resolving these variant gene families into large groups that should be further refined using other alignment free approaches such as networks. Network analysis, as has been previously used in classifying var gene (Bull et al., 2008) should be utilized in pir gene clustering in future analyses.

Recent studies have identified *rif* and *stevor* genes as additional mediators of rosetting and invasion in *P. falciparum* (Goel et al., 2015, Niang et al., 2014), and it is likely that they are also involved in mediating binding to endothelial cells. This study has identified dominantly expressed *rif* and *stevor* genes in parasites selected for binding to HBEC binding and rosetting. More targeted localization and functional studies can be designed to determine whether *pirs* play a direct role in binding phenotypes.

9. APPENDIX

Appendix Table 9.1. Old and new 3D7 rif gene IDs.

3D7 rif n=184

new gene ID	old gene ID	new gene ID	old gene ID	new gene ID	old gene ID
PF3D7_0100200	PFA0010c	PF3D7_0617700	PFF0855c	PF3D7_1040500	PF10_0398
PF3D7_0100400	PFA0020w	PF3D7_0631700	PFF1540w	PF3D7_1040600	PF10_0399
PF3D7_0100600	PFA0030c	PF3D7_0631800	PFF1545w	PF3D7_1040700	PF10_0400
PF3D7_0100800	PFA0040w	PF3D7_0632000	PFF1555w	PF3D7_1040800	PF10_0401
PF3D7_0100900	PFA0045c	PF3D7_0632100	PFF1560c	PF3D7_1040900	PF10_0402
PF3D7_0101000	PFA0050c	PF3D7_0632200	PFF1565c	PF3D7_1041000	PF10_0403
PF3D7_0101600	PFA0080c	PF3D7_0632300	PFF1570w	PF3D7_1041100	PF10_0404
PF3D7_0101900	PFA0095c	PF3D7_0632400	PFF1575w	PF3D7_1041200	PF10_0405
PF3D7_0114700	PFA0710c	PF3D7_0632600	PFF1585w	PF3D7_1100300	PF11_0009
PF3D7_0115200	PFA0740w	PF3D7_0632700	PFF1590w	PF3D7_1100400	PF11_0010
PF3D7_0115300	PFA0745w	PF3D7_0700200	MAL8P1.219	PF3D7_1100500	PF11_0011
PF3D7_0115600	PFA0760w	PF3D7_0700300	MAL8P1.218	PF3D7_1101100	PF11_0529
PF3D7_0200200	PFB0015c	PF3D7_0700500	MAL8P1.216	PF3D7_1101200	PF11_0020
PF3D7_0200500	PFB0030c	PF3D7_0701100	New rif	PF3D7_1101300	PF11_0021
PF3D7_0200600	PFB0035c	PF3D7_0701200	New rif	PF3D7_1101400	PF11_0022
PF3D7_0200700	PFB0040c	PF3D7_0701300	MAL7P1.200	PF3D7_1149800	PF11_0515
PF3D7_0201000	PFB0055c	PF3D7_0701800	PF07_0003	PF3D7_1150000	PF11_0517
PF3D7_0201200	PFB0060w	PF3D7_0711900	MAL7P1.43	PF3D7_1150100	PF11_0518
PF3D7_0222400	PFB1000w	PF3D7_0712200	MAL7P1.47	PF3D7_1150200	PF11_0519
PF3D7_0222500	PFB1005w	PF3D7_0712500	MAL7P1.52	PF3D7_1150300	PF11_0520
PF3D7_0222600	PFB1010w	PF3D7_0713000	MAL7P1.57	PF3D7_1200200	PFL0010c
PF3D7_0222700	PFB1015w	PF3D7_0731900	MAL7P1.179	PF3D7_1200300	PFL0015c
PF3D7_0223100	PFB1035w	PF3D7_0732200	PF07_0132	PF3D7_1200500	PFL0025c
PF3D7_0223200	PFB1040w	PF3D7_0732300	MAL7P1.184	PF3D7_1219200	PFL0933w
PF3D7_0223400	PFB1050w	PF3D7_0732400	PF07_0134	PF3D7_1240700	PFL1965w
PF3D7_0300200	PFC0010c	PF3D7_0732500	MAL7P1.185	PF3D7_1253600	PFL2580w
PF3D7_0300500	PFC0030c	PF3D7_0732700	PF07_0136	PF3D7_1253700	PFL2585c
PF3D7_0300700	PFC0035w	PF3D7_0732900	PF07_0138	PF3D7_1254000	PFL2605w
PF3D7_0300800	PFC0040w	PF3D7_0800400	PF08_0139	PF3D7_1254200	PFL2615w
PF3D7_0324400	PFC1095w	PF3D7_0800500	PF08_0138	PF3D7_1254400	PFL2625w
PF3D7_0324500	PFC1100w	PF3D7_0808800	PF08_0105	PF3D7_1254500	PFL2630w
PF3D7_0324800	PFC1115w	PF3D7_0808900	PF08_0104	PF3D7_1254700	PFL2640c
PF3D7_0400300	PFD0015c	PF3D7_0832100	MAL7P1.226	PF3D7_1254800	PFL2645c
PF3D7_0400500	PFD0025w	PF3D7_0832500	MAL7P1.222	PF3D7_1255000	PFL2655w
PF3D7_0400700	PFD0030c	PF3D7_0832800	MAL7P1.219	PF3D7_1255100	PFL2660w
PF3D7_0400900	PFD0040c	PF3D7_0833000	MAL7P1.217	PF3D7_1300200	MAL13P1.2
PF3D7_0401000	PFD0045c	PF3D7_0833100	MAL7P1.216	PF3D7_1300400	PF13_0004
PF3D7_0401200	PFD0050w	PF3D7_0833200	MAL7P1.215	PF3D7_1300500	PF13_0005
PF3D7_0401300	PFD0055w	PF3D7_0833400	MAL7P1.213	PF3D7_1300600	PF13_0006
PF3D7_0401400	PFD0060w	PF3D7_0900200	PFI0010c	PF3D7_1300700	MAL13P1.4
PF3D7_0401600	PFD0070c	PF3D7_0900300	PFI0015c	PF3D7_1301000	MAL13P1.8
PF3D7_0402500	PFD0120w	PF3D7_0900400	PFI0020w	PF3D7_1372600	MAL13P1.495
PF3D7_0402700	PFD0135c	PF3D7_0900500	PFI0025c	PF3D7_1372700	MAL13P1.500
PF3D7_0412600	PFD0620c	PF3D7_0900600	PFI0030c	PF3D7_1373000	MAL13P1.515
PF3D7_0413200	PFD0640c	PF3D7_0900700	PFI0035c	PF3D7_1373100	MAL13P1.520
PF3D7_0413300	PFD0645w	PF3D7_0901000	PFI0050c	PF3D7_1373300	MAL13P1.530
PF3D7_0421200	PFD1010w	PF3D7_0901100	PFI0055c	PF3D7_1373400	MAL13P1.535
PF3D7_0421500	PFD1020c	PF3D7_0901300	PFI0065w	PF3D7_1400200	PF14_0002
PF3D7_0425600	PFD1225w	PF3D7_0901400	PFI0070w	PF3D7_1400300	PF14_0003
PF3D7_0425700	PFD1230c	PF3D7_0901500	PFI0075w	PF3D7_1400400	PF14_0004
PF3D7_0425900	PFD1240w	PF3D7_0937300	PFI1805w	PF3D7_1400500	PF14_0005
PF3D7_0500300	PFE0015c	PF3D7_0937400	PFI1810w	PF3D7_1400600	PF14_0006
PF3D7_0500400	PFE0020c	PF3D7_0937500	PFI1815c	PF3D7_1400800	PF14_0008
PF3D7_0500500	PFE0025c	PF3D7_0937700	PFI1825w	PF3D7_1479400	PF14_0766
PF3D7_0500700	PFE0035c	PF3D7_1000200	PF10_0002	PF3D7_1479600	PF14_0768
PF3D7_0532900	PFE1630w	PF3D7_1000300	PF10_0003	PF3D7_1479700	PF14_0769
PF3D7_0533000	PFE1635w	PF3D7_1000400	PF10_0004	PF3D7_1479800	PF14_0770
PF3D7_0600300	PFF0015c	PF3D7_1000500	PF10_0005	PF3D7_1480000	PF14_0772
PF3D7_0600500	PFF0025w	PF3D7_1000600	PF10_0006		
PF3D7_0600700	PFF0035c	PF3D7_1040000	PF10_0393		
PF3D7_0600800	PFF0040c	PF3D7_1040100	PF10_0394		
PF3D7_0600900	PFF0045c	PF3D7_1040300	PF10_0396		
PF3D7_0617500	PFF0847w	PF3D7_1040400	PF10_0397		

Appendix Table 9.2. Old and new 3D7 stevor gene IDs.

3D7 stevor n=42	
new gene ID	old gene ID
PF3D7_0532800	PFE1627c
PF3D7_0200400	PFB0025c
PF3D7_0114600	PFA0705c
PF3D7_0115400	PFA0750w
PF3D7_0200900	PFB0050c
PF3D7_0300900	PFC0045w
PF3D7_0300400	PFC0025c
PF3D7_0324600	PFC1105w
PF3D7_0400800	PFD0035c
PF3D7_0401500	PFD0065w
PF3D7_0402600	PFD0125c
PF3D7_0425500	PFD1220c
PF3D7_0500600	PFE0030c
PF3D7_0900900	PFI0045c
PF3D7_0901600	PFI0080w
PF3D7_1000800	PF10_0009
PF3D7_1040200	PF10_0395
PF3D7_0832900	MAL7P1.218
PF3D7_1149900	PF11_0516
PF3D7_1300900	PF13_0009
PF3D7_1372500	MAL13P1.490
PF3D7_1372800	MAL13P1.505
PF3D7_0832600	MAL7P1.310
PF3D7_0832400	MAL7P1.223
PF3D7_0832000	MAL7P1.227
PF3D7_0700400	MAL8P1.217
PF3D7_0700700	MAL8P1.214
PF3D7_0102100	PFA0105w
PF3D7_0101800	PFA0090c
PF3D7_0732000	PF07_0130
PF3D7_0222800	PFB1020w
PF3D7_0221400	PFB0955w
PF3D7_0201300	PFB0065w
PF3D7_0631900	PFF1550w
PF3D7_0617600	PFF0850c
PF3D7_1254600	PFL2635w
PF3D7_1254300	PFL2620w
PF3D7_1254100	PFL2610w
PF3D7_1100700	PF11_0013
PF3D7_1479900	PF14_0771
PF3D7_1479500	PF14_0767
PF3D7_1400700	PF14_0007

Appendix Table 9.3. Conserved rif genes.

Group (members)	Lab Isolate	Gene ID	Old ID (3D7)	Type	Idnetitical nucleotides (% pairwise identity)	Patient isolate with ortholog	Gene ID of ortholog	Idnetitical nucleotides (% pairwise identity)	Group mean % pairwise identity	Ups sequence Identical nucleotides (% pairwise identity) for IT and 3D7
1 (21)	3D7	PF3D7_0808900	PF08_0104	A	Query	6816	PF8443_8_1_U073300	916/1000 (91%)	91.3	946/1010 (93%)
	IT	PFIT_1240200			921/999 (92%)	6816	PF8443_8_1_U547100	916/999 (91%)		
	HB3	PfHB3_BroadWash172400			912/999 (91%)	10936	PF8443_8_10_U548000	939/999 (93%)		
	IT	PFIT_bin08400			916/1005 (91%)	9626	PF8443_8_2_U143100	917/999 (91%)		
	DD2	PFDG_04323			928/999 (92%)	9626	PF8443_8_2_U513000	918/1002 (91%)		
						11019	PF8443_8_3_U164800	965/1020 (94%)		
						11019	PF8443_8_3_U477700	923/999 (92%)		
						9215	PF8443_8_4_U071400	876/961 (91%)		
						9215	PF8443_8_4_U195200	911/1001 (91%)		
						9775	PF8443_8_5_U165400	884/950 (93%)		
						9605	PF8443_8_6_U548500	923/1002 (92%)		
						9106	PF8443_8_7_U513600	908/999 (90%)		
						10668	PF8443_8_8_U293700	952/999 (95%)		
						11014	PF8443_8_9_U142300	951/999 (95%)		
						11014	PF8443_8_9_U162600	916/999 (91%)		
2 (20)	3D7	PF3D7_0412600 (s)	PFD0620c	A ps	Query	6816	PF8443_8_1_U561100	980/1023 (95%)	93.4	946/1042 (90%)
	IT	PFIT_0411200 (s)			975/1023 (95%)	10936	PF8443_8_10_U489500	986/1023 (96%)		
	3D7	PF3D7_0711900	MAL7p1.43		1013/1074 (94%)	10936	PF8443_8_10_U489700	985/1073 (91%)		
	IT	PFIT_0811700			851/934 (91%)	10936	PF8443_8_10_U489800	959/1036 (92%)		
	HB3	PfHB3_BroadWash144100			1019/1074 (94%)	9626	PF8443_8_2_U283400	885/939 (94%)		
	HB3	PfHB3_BroadWash054500			971/1053 (92%)	11019	PF8443_8_3_U116300	970/1023 (94%)		
						11019	PF8443_8_3_U546400	787/816 (96%)		
						9215	PF8443_8_4_U515500	909/959 (94%)		
						9215	PF8443_8_4_U515600	853/905 (94%)		
						9775	PF8443_8_5_U058900	825/859 (96%)		
						10668	PF8443_8_8_U508400	828/870 (95%)		
						11014	PF8443_8_9_U052000	928/970 (95%)		
						SA075	PFSAO75_071600	978/1074 (91%)		
						SA075	PFSAO75_137600	951/1023 (92%)		
3 (15)	3D7	PF3D7_0114700 (s)	PFA0710c	B3	Query	10668	PF8443_8_8_U549400	970/972 (99%)	98.9	992/1004 (98%)
	IT	PFIT_0113500 (s)			971/972 (99%)	6816	PF8443_8_1_U531100	970/972 (99%)		
	DD2	PFDG_03938			971/972 (99%)	11019	PF8443_8_3_U570000	970/972 (99%)		
	HB3	PfHB3_BroadWash537700			603/603 (100%)	9215	PF8443_8_4_U546500	968/972 (99%)		
	PRCDC	PRCDC_0014300			915/967 (94%)	SA075	PFSAO75_011200	967/973 (99%)		
						Brazilian	DQ265207	866/868 (99%)		
						10738	10739	803/803 (100%)		
						10936	PF8443_8_10_U493400	603/603 (100%)		
						9106	PF8443_8_7_U526900	603/603 (100%)		
						9605	PF8443_8_6_U262000	602/603 (99%)		
4 (14)	3D7	PF3D7_0713000 (s)	MAL7P1.57	A	Query	10936	PF8443_8_10_U132900	508/525 (96%)	97.4	946/1025 (92%)
	IT	PFIT_0711100 (s)			1021/1041 (98%)	10975	PF8443_8_12_U137000	924/963 (95%)		
	HB3	PfHB3_BroadWash144500			512/525 (97%)	9626	PF8443_8_2_U127100	1026/1041 (98%)		
						11019	PF8443_8_3_U123700	1010/1041 (97%)		
						9215	PF8443_8_4_U144300	1025/1041 (98%)		
						9775	PF8443_8_5_U139700	1009/1041 (96%)		
						9605	PF8443_8_6_U140200	1027/1041 (98%)		
						9106	PF8443_8_7_U060200	994/1041 (95%)		
						10668	PF8443_8_8_U135200	1022/1024 (99%)		
						11014	PF8443_8_9_U121800	1010/1041 (97%)		
						SA075	PFSAO75_137800	508/525 (96%)		
5 (14)	3D7	PF3D7_1101400	PF11_0022	B ps	Query	SA075	PFSAO75_128500	1038/1041 (99%)	95.5	843/1001 (84%)
	IT	PFIT_1100600			1011/1083 (93%)	SA075	PFSAO75_154900	1080/1083 (99%)		
	3D7	PF3D7_0631700	PFF1540w		1083/1083 (100%)	6816	PF8443_8_1_U557500	878/879 (99%)		
	3D7	PF3D7_0701300	MAL7P1.200		1083/1084 (99%)	11019	PF8443_8_3_U549100	870/875 (99%)		
	3D7	PF3D7_0222400			1002/1077 (93%)	9605	PF8443_8_6_U525800	831/879 (94%)		
	3D7	PF3D7_1040000			995/1077 (92%)	10936	PF8443_8_10_U495700	804/873 (92%)		
	DD2	PFDG_04006			872/891 (97%)	11019	PF8443_8_3_U533700	656/699 (93%)		
6 (14)	3D7	PF3D7_0402500 (s)	PFD0120w	A ps		9605	PF8443_8_6_U068900	982/990 (99%)	95.6	995/1004 (99%)
	IT	PFIT_0401200 (s)			970/990 (97%)	9215	PF8443_8_4_U069900	975/990 (98%)		
	HB3	PfHB3_BroadWash070900			923/1000 (92%)	10668	PF8443_8_8_U066200	971/990 (98%)		
						9106	PF8443_8_7_U55080	966/990 (97%)		
						10936	PF8443_8_10_U523100	964/990 (97%)		
						SA075	PFSAO75_052700	962/990 (97%)		
						11019	PF8443_8_3_U516200	923/1000 (92%)		
						Brazilian	DQ265340	808/815 (99%)		
						10975	PF8443_8_12_U068800	835/897 (93%		
						6816	PF8443_8_1_U071700	731/753 (97%)		
						9775	PF8443_8_5_U069200	692/706 (98%)		

7 (13)	3D7 IT PRCDC	PF3D7_1219200 (s) PFIT_1218900 (s) PRCDC_1218500 (s)	PFL0933w	A ps	Query 704/704 (100%) 688/704 (97%)	10936 10975 9626 11019 9215 9775 9106 10668 11014 SA075	PF8443_8_10_U495100 PF8443_8_12_U312700 PF8443_8_2_U300600 PF8443_8_3_U279800 PF8443_8_4_U327700 PF8443_8_5_U324600 PF8443_8_7_U276900 PF8443_8_8_U309400 PF8443_8_9_U278200 PFSAO75_318800	625/629 (99%) 701/704 (99%) 700/704 (99%) 640/644 (99%) 626/629 (99%) 701/704 (99%) 626/629 (99%) 625/629 (99%) 701/704 (99%) 701/704 (99%)	99.4	970/1030 (94%)
8 (13)	3D7 IT DD2 PRCDC	PF3D7_0413300 (s) PFIT_0411800 (s) PFDG_03424 PRCDC_0410800*	PFD0645w	A	Query 1017/1041 (97%) 1037/1041 (99%) 765/1049 (72%)*	6816 10975 9626 11019 9215 9775 10668 11014 SA075	PF8443_8_1_U061600 PF8443_8_12_U07020 PF8443_8_2_U053400 PF8443_8_3_U052600 PF8443_8_4_U059600 PF8443_8_5_U059100 PF8443_8_8_U056700 PF8443_8_9_U052300 PFSAO75_061500	1031/1041 (99%) 846/860 (98%) 1029/1041 (98%) 1015/1041 (97%) 945/1041 (90%) 1037/1041 (99%) 945/1041 (90%) 1037/1041 (99%) 1016/1041 (97%)	96	990/1010 (98%)
9 (13)	3D7 IT HB3 DD2 PRCDC	PF3D7_0401600 PFIT_0400300.2 PfHB3_BroadWash518600 PFDG_04771 PRCDC_0046400	PFD00700c	A2	Query 1029/1081 (95%) 1020/1065 (95%) 825/897 (91%) 839/974 (86%)*	6816 9215 11019 9605 9106 10936 8383 SA075	PF8443_8_1_U501500 PF8443_8_4_U026100 PF8443_8_3_U495100 PF8443_8_6_U553500 PF8443_8_7_U563300 PF8443_8_10_U484500 PF8443_8_11_U383500 PFSAO75_051800	963/1048 (91%) 1025/1050 (97%) 859/867 (99%) 864/871 (99%) 1030/1065 (96%) 1035/1044 (99%) 838/839 (99%) 863/871 (99%)	93.2	961/1038 (92%)
10 (13)	3D7 IT HB3	PF3D7_0421500 PFIT_0419700 PfHB3_BroadWash075100	PFD1020c	A ps	769/769 (100%) 771/786 (98%)	9215 6816 9605 11019 11014 10936 10668 10595 9775 10975	PF8443_8_4_U06700 PF8443_8_1_U068700 PF8443_8_6_U066300 PF8443_8_3_U059600 PF8443_8_9_U059200 PF8443_8_10_U065900 PF8443_8_8_U063400 595 PF8443_8_5_U066400 PF8443_8_12_U065600	769/769 (100%) 768/769 (99%) 767/769 (99%) 764/769 (99%) 764/769 (99%) 769/784 (98%) 761/769 (98%) 648/648 (100%) 586/587 (99%) 565/572 (98%)	99	999/1000 (99%)
11 (11)	3D7 IT DD2 HB3	PF3D7_0731900 (s) PFIT_0730900 (s) PFDG_02022 PfHB3_BroadWash532400	MAL7P1.179	B ps	Query 942/958 (98%) 590/601 (98%) 558/560 (99%)	SA075 10668 6816 10936 9605 9106 9215	PFSAO75_154300 PF8443_8_8_U526000 PF8443_8_1_U542300 PF8443_8_10_U545400 PF8443_8_6_U529600 PF8443_8_7_U471900 PF8443_8_4_U530800	956/957 (99%) 865/867 (99%) 831/867 (95%) 559/560 (99%) 559/560 (99%) 559/560 (99%) 559/560 (99%)	99.2	779/1086 (71%)
12 (10)	3D7 HB3 DD2	PF3D7_0532900 PfHB3_BroadWash536000 PFDG_02919	PFE1630w	B3	Query 956/957 (99%) 868/958 (90%)	9106 10668 9605 SA075 6816 10936 Brazilian	PF8443_8_7_U444000 PF8443_8_8_U484100 PF8443_8_6_U537200 PFSAO75_101100 PF8443_8_1_U511400 PF8443_8_10_U150700 DQ265354	956/957 (99%) 956/957 (99%) 948/957 (99%) 882/958 (92%) 868/958 (90%) 867/958 (90%) 745/791 (94%)	94.8	N/A
13 (10)	3D7 IT HB3 DD2	PF3D7_1372600 PFIT_0733200 PfHB3_BroadWash519800 PFDG_00493.1	MAL13P1.49	B3	Query 817/819 (99%) 727/727 (100%) 819/819 (100%)	9215 9106 9606 SA075 10936 6816	PF8443_8_4_U557600 PF8443_8_7_U551100 PF8443_8_6_U517200 PFSAO75_412900 PF8443_8_10_U549400 PF8443_8_1_U512500	768/772 (99%) 682/708 (96%) 761/769 (98%) 812/819 (99%) 768/772 (99%) 760/769 (98%)	98.8	992/1008 (98%)
14 (10)	3D7 IT HB3 PRCDC	PF3D7_0533000 (s) PFIT_0536700 (s) PfHB3_BroadWash536100 PRCDC_0531900	PFE1635w	A ps	Query 968/1029 (94%) 633/639 (99%) 906/1015 (89%)*	SA075 10936 6816 9605 9215 11019	PFSAO75_PN22 PF8443_8_10_U544600 PF8443_8_1_U511300 PF8443_8_6_U537300 PF8443_8_4_U512100 PF8443_8_3_U507000	884/891 (99%) 686/697 (98%) 633/639 (99%) 628/639 (98%) 596/617 (96%) 540/556 (97%)	97.5	955/1040 (91%)
15 (9)	3D7	PF3D7_0632300	PFF1570w	B	Query	10595 10668 10936 SA075 9605 Brazilian 10594 11019	595g_11 PF8443_8_8_U522100 PF8443_8_10_U490800 PFSAO75_127700 PF8443_8_6_U519500 DQ265189.1|Pf42_54|B 594g_28 PF8443_8_3_U564000	858/879 (97%) 847/878 (96%) 878/878 (100%) 877/878 (99%) 876/878 (99%) 875/878 (99%) 849/878 (96%) 868/878 (98%)	97.9	N/A

16 (9)	3D7 DD2	PF3D7_0808800 PFDG_04322	PF08_0105	A	Query 1008/1114 (90%)	9215 10668 6816 9626 11014 6816 11019	PF8443_8_4_U171000 PF8443_8_8_U293600 PF8443_8_1_U560100 PF8443_8_2_U143200 PF8443_8_9_U502600 PF8443_8_1_U547200 PF8443_8_3_U568200	998/1070 (93%) 1022/1113 (91%) 1020/1113 (91%) 879/891 (98%) 1014/1116 (90%) 964/1038 (92%) 923/976 (94%)	91.4	N/A
17 (8)	3D7	PF3D7_0413200	PFD0640c	A	Query	6816 9626 11019 9775 9605 10668 SA075	PF8443_8_1_U061500 PF8443_8_2_U053300 PF8443_8_3_U052500 PF8443_8_5_U059000 PF8443_8_6_U059100 PF8443_8_8_U056600 PFSAO75_061400	1050/1134 (92%) 1131/1131 (100%) 1103/1131 (97%) 889/939 (94%) 1102/1131 (97% 1105/1131 (97%) 1021/1131 (90%)	94.1	N/A
18 (8)	3D7 3D7 3D7	PF3D7_0425600 PF3D7_0700300 PF3D7_1100500	PFD1225w MAL8P1.218 PF11_0011	A ps	Query 852/855 (99%) 865/915 (94%)	10668 Brazilian 10936 SA075 9106	PF8443_8_8_U553400 DQ265170.1 PF8443_8_10_U561800 PFSAO75_128300 PF8443_8_7_U447200	784/788 (99%) 781/786 (99%) 783/788 (99%) 784/788 (99%) 782/788 (99%)	98.5	
19 (8)	3D7 IT HB3	PF3D7_1240700 PFIT_1240800 PfHB3_BroadWash054500	PFL1965w	A ps	Query 773/808 (95%) 759/809 (93%)	9215 10936 10936 SA075 9626	PF8443_8_4_U561700 PF8443_8_10_U294000 PF8443_8_10_U489800 PFSAO75_071600 PF8443_8_2_U283400	710/750 (94%) 699/731 (95%) 740/809 (91%) 749/832 (90%) 675/704 (95%)	92.8	941/1015 (92%)
20 (7)	3D7	PF3D7_0500300	PFE0015c	B ps	Query	9106 9215 10594 10668 9215 10936	PF8443_8_7_U458100 PF8443_8_4_U523700 594_1_2 PF8443_8_8_U097200 PF8443_8_4_U506700 PF8443_8_10_U548700	847/847 (100%) 463/463 (100%) 839/839 (100%) 374/389 (96%) 889/939 (94%) 931/931 (100%)	97.7	N/A
21 (7)	IT DD2	PFIT_0616600 PFDG_02423		A	Query 663/725 (91%)	10936 9215 9775 9106 SA075	PF8443_8_10_U110500 PF8443_8_4_U134400 PF8443_8_5_U130900 PF8443_8_7_U112500 PFSAO75_PN18	1048/1148 (91%) 1137/1147 (99%) 1137/1147 (99%) 1049/1148 (91%) 1059/1069 (99%	94	N/A
22 (7)	3D7 IT HB3	PF3D7_1253700 PFIT_0801500 PfHB3_BroadWash515900	PFL2585c	A	Query 1125/1128 (99%) 1048/1050 (99%)	Brazilian 9605 10668 9215	DQ265365 PF8443_8_6_U541700 PF8443_8_8_U487800 PF8443_8_4_U509500	875/923 (94%) 1049/1050 (99% 1049/1050 (99%) 1048/1050 (99%)	95.6	588/1101 (53%)
23 (7)	3D7 3D7 HB3 PRCDC	PF3D7_0901100 PF3D7_1301000 PfHB3_BroadWash534600 PRCDC_0006200	PFI0055c MAL13P1.8	A	Query 1048/1052 (99%) 702/703 (99%) 854/1052 (81%)*	6816 9215 11019	PF8443_8_1_U546400 PF8443_8_4_U550500 PF8443_8_3_U511600	1050/1053 (99%) 714/715 (99% 702/703 (99%)	99.8	N/A
24 (7)	3D7 IT	PF3D7_0402700 (s) PFIT_0401400 (s)	PFD0135c	A ps	Query 981/990 (99%)	SA075 9106 10975 10936 6816	PFSAO75_052900 PF8443_8_7_U471700 PF8443_8_12_U068900 PF8443_8_10_U522900 PF8443_8_1_U071900	941/1007 (93%) 789/796 (99%) 734/795 (92%) 734/795 (92%) 734/795 (92%)	92.5	1000/1000 (100%)
25 (6)	3D7 IT DD2	PF3D7_0712500 PFIT_0411700 PFDG_00027	MAL7P1.52	A	Query 995/997 (99%) 1106/1139 (97%)	9775 9626 6816	PF8443_8_5_U139500 PF8443_8_2_U126900 PF8443_8_1_U141200	1107/1137 (97%) 1089/1137 (95%) 1043/1065 (97%)	98.4	690/1113 (61%)
26 (5)	3D7 HB3	PF3D7_1000600 PfHB3_BroadWash262300	PF10_0006	B1	Query 1061/1086 (97%)	6816 SA075 9106	PF8443_8_1_U554200 PFSAO75_220500 PF8443_8_7_U502300	1031/1087 (94%) 1025/1087 (94%) 936/957 (97%)	96.1	N/A
27 (4)	3D7	PF3D7_0617500	PFF0847w	A	Query	9106 SA075 9775	PF8443_8_7_U100400 PFSAO75_071100 PF8443_8_5_U530200	886/892 (99%) 582/586 (99%) 582/586 (99%)	99.5	N/A
28 (4)	3D7	PF3D7_0600900	PFF0045c	B	Query	10975 11019 6816	PF8443_8_10_U566400 PF8443_8_3_U548900 PF8443_8_1_U561800	832/845 (98%) 884/970 (91%) 720/771 (93%)	93	N/A
29 (3)	3D7	PF3D7_0617700	PFF0855c	A	Query	9106 SA075	PF8443_8_7_U100500 PFSAO75_071300	906/1003 (90%) 717/752 (95%)	92.6	N/A
30 (3)	3D7	PF3D7_0712200	MAL7P1.47	A	Query	9775 11014	PF8443_8_5_U139300 PF8443_8_9_U435200	1014/1017 (99%) 944/1018 (92%)	94.9	N/A
31 (3)	3D7 DD2	PF3D7_0900300 PFDG_00601	PFI0015c	B2	Query 997/1081 (92%)	9106	PF8443_8_7_U511400	947/1011 (93%)		
32 (3)	3D7 IT	PF3D7_0832100 (s) PFIT_0835100 (s)	MAL7P1.226	A ps	895/903 (99%)	SA075	PFSAO75_184200	942/950 (99%)	99.3	644/1354 (47%)!
33 (2)	3D7	PF3D7_0222700	PFB1015w	B	Query	6816	PF8443_8_1_U550200	845/845 (100%)	100	N/A
34 (2)	3D7	PF3D7_0833400	MAL7P1.213	A	Query	SA075	PFSAO75_PN57	798/878 (90%)	90.9	N/A
35 (2)	3D7 IT	PF3D7_1253600 PFIT_0801600	PFL2580w	A ps	770/774 (99%)				99.5	584/1140 (51%)

Appendix Table 9.4. Conseved stevor genes.

		Laboratory isolates				Wild isolates				
Group embers)	Lab Isolate	Gene ID	Old ID (3D7)	Type	Idnetitical nucleotides (% pairwise identity)	Patient isolate with ortholog	Gene ID of ortholog	Idnetitical nucleotides (% pairwise identity)	Group mean % pairwise identity	Ups sequence Identical nucleotides (% pairwise identity) for IT and 3D7
1 (12)	3D7	PF3D7_0402600 (s)	PFD0125c	B	Query	SA075	PFSAO75_052800	902/907 (99%)	99.7	984/1016 (96%)
	IT	PFIT_0401300 (s)			905/906 (99%)	9605	PF8443_8_6_U069000	859/859 (100%)		
	DD2	PFDG_01541			897/907 (98%)	11019	PF8443_8_3_U516100	594/594 (100%)		
	HB3	PfHB3_BroadWash071000			594/594 (100%)	9775	PF8443_8_5_U069100	594/594 (100%)		
						10936	PF8443_8_10_U523000	594/594 (100%)		
						9215	PF8443_8_4_U070000	594/594 (100%)		
						6816	PF8443_8_1_U071800	593/594 (99%		
						9106	PF8443_8_7_U550700	593/594 (99%)		
2 (11)	3D7	PF3D7_0201300	PFB0065w	A	Query	SA075	PFSAO75_011900	915/918 (99%)	98.7	N/A
						11019	PF8443_8_3_U441000	896/897 (99%)		
						10936	PF8443_8_10_U343600	896/897 (99%)		
						6816	PF8443_8_1_U528400	895/897 (99%)		
						10668	PF8443_8_8_U560900	892/897 (99%)		
						Brazilian	DQ265680	716/725 (98%)		
						Brazilian	DQ265679	716/725 (98%)		
						Brazilian	DQ265676	715/725 (98%)		
						Brazilian	DQ265677	714/725 (98%)		
						Brazilian	DQ26568	696/717 (97%)		
3 (11)	3D7	PF3D7_0832000 (s)	MAL7P1.227	B	Query	11019	PF8443_8_3_U527500	894/939 (95%)	95.9	995/1004 (99%)
	IT	PFIT_0835000 (s)			877/939 (93%)	10668	PF8443_8_8_U498100	891/939 (94%)		
	HB3	PfHB3_BroadWash548600			869/939 (92%)	9605	PF8443_8_6_U188700	879/912 (96%)		
						9215	PF8443_8_4_U529500	879/939 (93%)		
						9775	PF8443_8_5_U188600	878/948 (92%)		
						6816	PF8443_8_1_U187900	869/939 (92%)		
						SA075	PFSAO75_184100	868/948 (91%)		
						9106	PF8443_8_7_U438100	749/820 (91%)		
4 (11)	3D7	PF3D7_1000800 (s)	PF10_0009	B ps	Query	9605	PF8443_8_6_U552400	876/879 (99%)	98.4	912/1022 (89%)
	IT	PFIT_1000100 (s)			823/874 (94%)	SA075	PFSAO75_220700	876/879 (99%)		
	HB3	PfHB3_BroadWash262600			872/879 (99%)	9215	PF8443_8_4_U265400	875/879 (99%)		
	DD2	PFDG_0102			872/879 (99%)	10936	PF8443_8_10_U546500	874/879 (99%)		
						11019	PF8443_8_3_U439500	699/702 (99%)		
						6816	PF8443_8_1_U523900	566/567 (99%		
						10668	PF8443_8_8_U556300	565/567 (99%)		
5 (10)	3D7	PF3D7_0102100 (s)	PFA0105w	A ps	Query	9106	PF8443_8_7_U55480	494/495 (99%	98.7	989/1010 (97%)
	IT	PFIT_0100600 (s)			833/846 (98%)	9215	PF8443_8_4_U537400	494/495 (99%)		
	HB3	PfHB3_BroadWash529200			483/495 (97%)	10975	PF8443_8_12_U009100	483/495 (97%)		
						10936	PF8443_8_10_U524300	483/495 (97%)		
						10668	PF8443_8_8_U489700	483/495 (97%)		
						6816	PF8443_8_1_U009600	483/495 (97%)		
						11019	PF8443_8_3_U497500	483/495 (97%)		
6 (6)	3D7	PF3D7_0400800	PFD0035c	A	Query	6816	PF8443_8_1_U514100	848/849 (99%	96	N/A
						10936	PF8443_8_10_U551900	822/871 (94%)		
						9215	PF8443_8_4_U503300	810/874 (92%)		
						11019	PF8443_8_3_U487900	707/720 (98%)		
						Brazilian	DQ265674	668/677 (98%)		
7 (6)	3D7	PF3D7_1400700	PF14_0007	A	Query	SA075	PFSAO75_258600	819/882 (92%)	93.3	N/A
						6816	PF8443_8_1_U548900	817/882 (92%)		
						9106	PF8443_8_7_U486000	810/882 (91%)		
						Brazilian	DQ265669	680/685 (99%)		
						Brazilian	DQ265668	681/689 (98%)		
8 (6)	3D7	PF3D7_0221400 (s)	PFB0955w	B	Query	SA075	PFSAO75_029900	917/924 (99%)	99.4	999/1000 (99%)
	IT	PFIT_0220800 (s)			923/924 (99%)	9106	PF8443_8_7_U454500	856/857 (99%)		
						10936	PF8443_8_10_U552700	855/857 (99%)		
						Brazilian	DQ265703	717/726 (98%)		
9 (6)	3D7	PF3D7_0700700	MAL8P1.214	B ps	Query	10668	PF8443_8_8_U550900	492/492 (100%)	97.9	N/A
	3D7	PF3D7_1100700	PF11_0013	B ps	822/864 (95%)	9605	PF8443_8_6_U503000	492/492 (100%)		
						11019	PF8443_8_3_U567200	491/492 (99%)		
						9215	PF8443_8_4_U551500	462/462 (100%)		
10 (6)	3D7	PF3D7_0222800	PFB1020w	A	Query	10936	PF8443_8_10_U521200	823/849 (96%)	90.2	604/1116 (54%)
	IT	PFIT_0700300			829/915 (90%)	11019	PF8443_8_3_U525000	843/927 (90%)		
	PRCDC	PRCDC_0045000			812/915 (88%)*	SA075	PFSAO75_128400	831/915 (90%)		

l1 (6)	3D7	PF3D7_0732000	PF07_0130	A	Query	SA075	PFSAO75_348400	827/891 (92%)	90.3	543/1191 (45%) !
	IT	PFIT_0900900			815/894 (91%)	9106	PF8443_8_7_U561400	821/895 (91%)		
	IT	PFIT_0900800			815/894 (91%)	9106	PF8443_8_7_U472600	811/890 (91%)		
12 (5)	3D7	PF3D7_0425500	PFD1220c	B	Query	6816	PF8443_8_1_U530000	912/915 (99%)	99.5	N/A
						9106	PF8443_8_7_U451500	912/915 (99%)		
						9605	PF8443_8_6_U542100	865/867 (99%)		
						Brazilian	DQ265704	728/736 (98%)		
13 (5)	3D7	PF3D7_0114600 (s)	PFA0705c	A ps	Query	SA075	PFSAO75_011100	924/924 (100%)	98.5	998/1002 (99%)
	IT	PFIT_0113400 (s)			924/924 (100%)					
	DD2	PFDG_03937			852/864 (98%)					
	PRCDC	PRCDC_0058900			678/700 (96%)					
14 (5)	3D7	PF3D7_0200900	PFB0050c	A ps	Query	10668	PF8443_8_8_U537100	853/912 (93%)	92.2	N/A
						10668	PF8443_8_8_U535700	819/900 (91%)		
						9106	PF8443_8_7_U543200	620/624 (99%)		
						SA075	PFSAO75_154400	820/909 (90%)		
15 (5)	3D7	PF3D7_1254300	PFL2620w	A	Query	10668	PF8443_8_8_U540800	891/909 (98%)	92	N/A
	3D7	PF3D7_1254600	PFL2635w	A	856/909 (94%)	SA075	PFSAO75_154400	833/909 (91%)		
						10668	PF8443_8_8_U535700	829/909 (91%)		
16 (4)	3D7	PF3D7_0617600	PFF0850c	A	Query	9106	PF8443_8_7_U112400	909/918 (99%)	98.3	N/A
						SA075	PFSAO75_071200	903/909 (99%)		
						9775	PF8443_8_5_U530300	905/925 (97%)		
17 (4)	3D7	PF3D7_1300900	PF13_0009	A	Query	9215	PF8443_8_4_U550600	923/923 (100%)	95.7	594/1102 (53%)
	HB3	PfHB3_BroadWash534700			923/927 (99%)					
	IT	PFIT_0731100			854/931 (91%)					
18 (4)	3D7	PF3D7_1372800	MAL13P1.505	A	Query	9605	PF8443_8_6_U544600	856/899 (95%)	93.5	N/A
	HB3	PfHB3_BroadWash549600			876/899 (97%)					
	HB3	PfHB3_BroadWash528000			806/887 (90%					
19 (4)	3D7	PF3D7_1372500 (s)	MAL13P1.490	B ps	Query	SA075	PFSAO75_412800	904/906 (99%)	99.3	998/1001 (99%)
	IT	PFIT_0733100 (s)			850/858 (99%)	SA075	PFSAO75_348200	901/906 (99%)		
20 (4)	3D7	PF3D7_1479500	PF14_0767	A	Query	9215	PF8443_8_4_U516100	833/915 (91%)	91.4	N/A
	HB3	PfHB3_BroadWash531600			825/902 (91%)	10668	PF8443_8_8_U553100	819/902 (90%)		
21 (4)	3D7	PF3D7_0401500 (s)	PFD0065w	A ps	Query	SA075	PFSAO75_051700	927/933 (99%)		978/1017 (96%)
	IT	PFIT_0400200 (s)			932/933 (99%)	Brazilian	DQ265618	691/736 (93%)		
22 (3)	3D7	PF3D7_0200400	PFB0025c	A	Query	9605	PF8443_8_6_U56000	882/915 (96%)	94.8	N/A
	DD2	PFDG_01019			849/894 (94%)					
23 (3)	3D7	PF3D7_0500600 (s)	PFE0030c	B	Query	SA075	PFSAO75_073300	923/924 (99%)	99.6	994/1006 (98%)
	IT	PFIT_0500700 (s)			919/924 (99%)					
24 (3)	3D7	PF3D7_0532800	PFE1627c	A ps	Query	SA075	PFSAO75_101000	700/705 (99%)		
	PRCDC	PRCDC_0531800			708/771 (91%)					
25 (3)	3D7	PF3D7_0631900	PFF1550w	A	Query	9106	PF8443_8_7_U436400	911/912 (99%)	95.1	N/A
						SA075	PFSAO75_493100	845/912 (92%)		
26 (3)	3D7	PF3D7_0700400	MAL8P1.217	A	Query	11019	PF8443_8_3_U562300	882/903 (97%)	96	N/A
						9106	PF8443_8_7_U541900	850/903 (94%)		
27 (2)	3D7	PF3D7_0300400	PFC0025c						90.3	N/A
	HB3	PfHB3_BroadWash162000			805/891 (90%)					
28 (2)	3D7	PF3D7_0832400	MAL7P1.223	A	Query	Brazilian	DQ265697	712/722 (98%)	98.6	N/A
	3D7	PF3D7_0900900	PFI0045c	A	Query	SA075	PFSAO75_PN63	758/810 (93%)	93.6	N/A
29 (2)	3D7	PF3D7_1149900	PF11_0516	A	Query	SA075	PFSAO75_001200	853/927 (92%)	92	N/A

Yellow highlights indicate the centrally located rifs.

The reference gene was used to search for orthologs using BLAST search in Geneious (version 6.1.5).

s indicates syntenic orthologs

** non-orthologs because of the low pairwise sequence identity suggesting*

Appendix Table 9.5. Parasite IDC time points, time in culture, staging by giemsa- stained slide, and for ex vivo patient isolates, estimated age in hours by micorarray.

Laboratory isolates			
Isolate	Time point	Hours in culture	IDC stage by slide (in cases of multiple stages the dominant stage is highlighted in bold)
3D7	T1	0	**Early ring**/ mid ring
	T2	12	**Late ring**/ early trophozoite
	T3	18	**Early trophozite**/ mid trophozoite/ late ring
	T4	24	**Mid trophozoite**/ late trophozoite
	T5	30	**late trophozoite**
	T6	42	**Schizont**/ late trophozoite/ early ring
IT	T1	0	**Early ring**
	T2	12	**Mid ring**/late ring
	T3	18	**Early trophozoite**/ late ring
	T4	24	**Mid trophozoite**/ early trophozoite
	T5	30	**Late trophozoite**
	T6	40	**Schizonts**/ late trophozoite
HB3	T1	0	**Early ring**
	T2	12	**Late ring**/ early trophozoite
	T3	18	**Early trophozoite**
	T4	24	**Mid trophozoite**
	T5	30	**Late trophozoite**
	T6	42	**Schizont**/ late trophozoite/ early ring
Laboratory-adapted Kilifi isolates			
Isolate	Time point	Hours in culture	IDC stage by slide (in cases of multiple stages the dominant stage is highlighted in bold)
8383	T1	0	Early to mid ring
	T2	6	Late ring/ early trophozoite
	T3	18	Mid/ late trophozoite
	T4	30	Schizont
9106	T1	0	Early/ mid ring
	T2	8	Late ring/ early trophozoite
	T3	18	Mid/ late trophozoite
	T4	26	Mid/ **late trophozoite/Sschizont**
9106HBEC	T1	0	Early ring
	T2	13	Late ring/ early trophozoite
	T3	25	Early/ **mid trophozoite**
	T4	44	**Schizont/** early ring
9215	T1	0	Early/ mid ring
	T2	10	**Late ring/** early trophozoite
	T3	19	**Mid**/ late trophozoite
	T4	30	**Late trophozoite**/ schizont
10668 R+3	T1	0	Mid ring
	T2	6	Late ring/ early trophozoite
	T3	20	Mid trophozoite
	T4	31	**Late trophozite/ schizont/** early ring
10668HBEC	T1	0	Early ring
	T2	20	Late ring/ early trophozoite
	T3	25	Early/ mid trophozoite
	T4	32	**Schizont/** early ring
10668 CU	T1	0	Early ring
	T2	20	Late ring/ early trophozoite
	T3	30	Early/ mid trophozoite
	T4	40	**Schizont/** early ring
SA075c R+	T1	0	Mid ring
	T2	10	Late ring/ early trophozoite
	T3	24	Mid/ late trophozoite
	T4	31	Schizont
SA075c R-	T1	0	Early to mid ring
	T2	22	Late ring/ early trophozoite
	T3	29	Mid trophozoite
	T4	48	Schizont

Ex-vivo field isolates (MJM's study) QC based on DD2								
Isolate	Timepoint	Time in culture	IDC stage by slide (in cases of multiple stages the dominant stage is highlighted in bold)	Array Estimated age (hrs)	*pir* expression by capillary Sanger sequencing (chapter 4)	Stage label on piecharts in chapter 4	*pir* expression by 454 sequencing (chapter 6)	Stage label on piecharts in chapter 6
10727	T1	0	Early ring	N/A			rif & stevor	R
	T2	10	N/A	N/A				
	T3	20	Mid ring	12			rif & stevor	
	T4	30	Early troph	14				
	T5	40	Mid troph	22			rif & stevor	T
	T6	50	late troph	N/A				
	T7	60	Schizont	32			rif & stevor	S
10761	T1	0	Early ring	N/A	rif & stevor	R		
	T2	10	Mid ring	12			rif	R
	T3	20	Mid ring	16	rif & stevor	LR-ET	rif	
	T4	30	Early troph	22			stevor	T
	T5	40	Late troph	26	rif	T	rif	T
	T6	50	Schizont	34			rif & stevor	S
	T7	60	Early ring	N/A	rif & stevor	S	rif	
10747	T1	0	Early ring	N/A				
	T2	10	Early ring	N/A				
	T3	20	N/A	N/A			rif & stevor	R
	T4	30	Mid ring	N/A				
	T5	40	Early troph	N/A			rif & stevor	LR-ET
	T6	50	N/A	N/A				
	T7	60	N/A	N/A			rif & stevor	S
10735	T1	0	Mid ring	N/A			rif	R
	T2	10	**Late ring**/ mid ring	N/A	rif	R		
	T3	20	Late ring/ **early trophozoite**	N/A	rif	LR-ET	rif	LR-ET
	T4	30	Early trophozoite	N/A	rif		rif	
	T5	40	Mid trophozoite	N/A	rif	T		
	T6	50	Mid trophozoite	N/A	rif			
	T7	60	**Late trophozoite**/ schizont	N/A				
10739	T1	0	**Mid ring**	N/A	rif & stevor	R	rif	R
	T2	10	**Late ring**/mid ring	N/A	rif		rif	
	T3	20	Late ring/**early trophozoite**	12	rif & stevor	LR-ET		
	T4	30	**Early trophozoite**/ mid trophozoite	N/A	rif	T		
	T5	40	Mid trophozoite	N/A	rif & stevor		rif	T
	T6	50	Late trophozoite	N/A	rif	S	rif	
	T7	60	**Late trophozoite**/ schizont	N/A	rif & stevor			
10814	T1	0	N/A	10	rif & stevor	ER		
	T2	10	N/A	12			rif & stevor	R
	T3	20	Mid ring	16	rif & stevor	LR-ET		
	T4	30	Mid ring	20			rif & stevor	PR
	T5	40	N/A	26	rif & stevor	T		
	T6	50	N/A	38				
	T7	60	Early troph	46	rif & stevor	S	rif & stevor	T
10594	T1	0	Mid ring	12	rif		rif	
	T2	10	**Late ring**/early trophozoite	12				
	T3	20	Mid trophozoite	18	rif		rif	
	T4	30	Late trophozoite	22	rif			
	T5	40	**Late trophozoite**/ early schizont	34				
	T6	50	**Schizont**/early ring	42	rif		rif	
	T7	60	**Late schizont**/ early ring	46				
10595	T1	0	Mid ring	10	rif		rif	
	T2	10	**Late ring**/early trophozoite	12	rif		rif	
	T3	20	mid trophozoite	20	rif		rif	
	T4	30	Late trophozoite	16	rif		rif	
	T5	40	**Late trophozoite**/ early schizont	30				
	T6	50	Schizont	40	rif		rif	
	T7	60	**Schizont**/ early ring	46				

Green highlight indcates overlapping time points where data were pooled for analysis by averaging the read counts of the two stages.

The stages in bold represent the dominant IDC stage in a time point with mixed stages.

Appendix Table 9.6. *Isolates used in the pir expression profiling study.*

Parasite isolate	Clinical phenotype of original isolate.	Parasite lines studied for *pir* expression during IDC	IDC stages of RNA sampling	Donor blood group	Rosetting frequency	No. of genotypes by MSP2 genotyping			Whole genome sequence	Partial pir assembly		Amplicon sequencing method	
						FC27 clones	IC clones	Total		No. of rif	No. of stevor	Capillary sequencing	454 sequencing
3D7	N/A	3D7	6 timepoints, ER, LR, ET, MT, LT, S of the IDC of the laboratory isolates	na	nd		1	1	Y	184 (complete)	42 (complete)	Y [r]	Y [r, s]
IT	N/A	IT		na	nd		1	1	Y	138	43	Y [r]	Y [r, s]
HB3	N/A	HB3		na	nd			1	Y	142	35	-	Y [r, s]
8383	NS	8383A	2 timepoints, *ex vivo* ring & trophozoite	B	nd	1		1	N	115§	19§	-	Y [r, s]
		8383_CU	4 timepoints, R, P, T and S of the Culture-adapted line	na	nd	1		1	Y[a]	17* (112§)	3* (20§)	-	Y [r, s]
9106	IC&RD	9106A	2 timepoints, *ex vivo* ring & trophozoite	B	9%	1		1	N			-	Y [r, s]
		9106_CU	4 timepoints, R, P, T and S of the Culture-adapted line	na	nd	1		1	Y	155	42	-	Y [r, s]
		9106_HB	4 timepoints, R, P, T and S of the Culture-adapted and HBEC-selected line	na	nd	1		1	N			-	Y [r, s]
9215	RD	9215A	2 timepoints, *ex vivo* ring & trophozoite	A	86.20%	nd	nd	nd	N			-	Y [r, s]
		9215_CU	4 timepoints, R, P, T and S of the Culture-adapted line	na	nd		2	2	Y	168	39	-	Y [r, s]
10668	IC&RD	10668A	2 timepoints, *ex vivo* ring & trophozoite	A	6.80%		2	2	N			-	Y [r, s]
		10668_CU	4 timepoints, R, P, T and S of the Culture-adapted line	na	4.30%		2	2	Y	179	40	-	Y [r, s]
		10668_HB	4 timepoints, R, P, T and S of the Culture-adapted and HBEC-selected line	na	nd		2	2	N				Y [r, s]
SA075	SMA	SA075_R+ (clone)	4 timepoints, R, P, T and S of the Culture-adapted and rosetting clone	nd	52%		2	2	N	95#*	31#*	Y [r, s]	Y [r, s]
		SA075_R- (uncloned fraction)	4 timepoints, R, P, T and S of the Culture-adapted rosetting negative uncloned fraction.	nd	0%		2	2	Y			Y [r, s]	Y [r, s]
10594	NS	10594	7 timepoints of *ex vivo* wild isolates (per isolate parasite stage in Appendix table 1)	O	nd	1	2	3	N			Y [r]	-
10595	NS	10595		O	nd	3	1	4	N			Y [r]	-
10735	NS	10735		O	nd	nd	nd	nd	N	124§	27§	Y [r]	Y [r]
10739	NS	10739		O	32.90%	4	0	4	N	137§	30§	Y [r, s]	Y [r]
10727	NS	10727		A	nd	0	1	1	N	193§	44§	-	Y [r, s]
10747	IC	10747		A	78.30%	3	1	4	N	241§	53§	-	Y [r, s]
10761	MM	10761		O	nd	1	1	2	N	206§	27§	Y [r, s]	Y [r, s]
10814	MM	10814		A	nd	1	1	2	N	355§	56§	Y [r, s]	Y [r, s]

**: Noticeably incomplete pir repertoire*

#: Additional 55 rifs and 9 stevors annotated in house

§: pir partial repertoires from de novo assembly of 454 amplicon sequencing reads

Appendix Table 9.7. List of primers used.

#	Primer Name	Sequence (space after each set of 3)	Procedure
1	Albrecht_R	5' - CAT AAA TGT TTC TTG CAY TCA TG	454 sequencing
2	Stevor-primer3_F_72	5' - WDH AGA ACC NAT GTC AAC G	454 sequencing
3	Lavazec_L	5' - CAA AAG GAA GAG ATA AGT AT	PCR for cloning and capillary sequencing
4	Lavazec_R	5' - GTT TCT TGC ATT CAT GTT TCC	PCR for cloning and capillary sequencing
5	Kyes-rifF1	5' - CRT CAC GAK TGT TAA GCG	PCR for cloning and capillary sequencing/454 sequencing
6	Kyes-rifF2	5' - CGA RYG TGA ATT GTA TGC	PCR for cloning and capillary sequencing/454 sequencing
7	Kyes-rifF3	5' - CYA CYA GRT TAT TAT GCG	PCR for cloning and capillary sequencing/454 sequencing
8	Kyes-rifRev	5' - CTT CAW ATT RTT WTT TYK DCG ATA ACG	PCR for cloning and capillary sequencing
9	Kyes-rifR1	5' - CTT CAW TTT RTT WTT TYK DCG ATA ACG	PCR for cloning and capillary sequencing
10	Kyes-rifR1-VT	5' - CTT CAW TTT YTT WTT TYK DCK ATA ACG	PCR for cloning and capillary sequencing/454 sequencing
11	Kyes-rifR1-T7	5' - cag aga tgc ata ata cga ctc act ata ggg CTT CAW TTT RTT WTT TYK DCG ATA ACG	rif northern probe
12	Kyes-rifF1-T7	5' - cag aga tgc ata ata cga ctc act ata ggg CRT CAC GAK TGT TAA GCG	rif northern probe
13	Kyes-rifF2-T7	5' - cag aga tgc ata ata cga ctc act ata ggg CGA RYG TGA ATT GTA TGC	rif northern probe
14	Kyes-rifF3-T7	5' - cag aga tgc ata ata cga ctc act ata ggg CYA CYA GRT TAT TAT GCG	rif northern probe
15	Kyes-rifR1-VT-T7	5' - cag aga tgc ata ata cga ctc act ata ggg CTT CAW TTT YTT WTT TYK DCK ATA ACG	rif northern probe
16	DBLaBR-T7	5' - cag aga tgc ata ata cga ctc act ata ggg GCC CAT TCS TCG AAC CA	var northern probe
17	DBLaAF-T7	5' - cag aga tgc ata ata cga ctc act ata ggg GCA CGM AGT TTY GC	var northern probe
18	MSP1F	5' - CTA GAA GCT TTA GAA GAT GCA GTA TTG	msp1 genotyping, notrhern probe
20	MSP1 R	5' - CTT AAA TAG TAT TCT AAT TCA AGT GGA TCA	msp1 genotyping
21	MSP1RKimura	5' - ATC CAT CAA TTA AAT ATT TGA AAC C	msp1 northern probe
22	MSP1RKimura-T7	5' - cag aga tgc ata ata cga ctc act ata ggg ATC CAT CAA TTA AAT ATT TGA AAC C	msp1 northern probe
23	msp2 F	5' - ATG AAG GTA ATT AAA ACA TTG TCT ATT ATA	msp2 genotyping
24	msp2 R	5' - CTT TGT TAC CAT CGG TAC ATT CTT	msp2 genotyping
25	Var-C1F	5' - AAA AAA CAA AAT CAT CAG TAG GAA ATT TAT T(C/T)C	var northern probe
26	Var-C2R	5' - TAT CCC ATA AAT CTG C(A/T)A T(A/T)G G(A/G)T A	var northern probe
27	Var-C1F-T7	5' - cag aga tgc ata ata cga ctc act ata ggg AAA AAA CAA AAT CAT CAG TAG GAA ATT TAT T(C/T)C	var northern probe
28	Var-C2R-T7	5' - cag aga tgc ata ata cga ctc act ata ggg TAT CCC ATA AAT CTG C(A/T)A T(A/T)G G(A/G)T A	var northern probe
29	calmodulinF	5' - TGGCAGACAAGTTAACAGAAGAACA	calmodulin northern probe
30	calmodulinR	5' - TGGCTATCATCATTTTAACAAACTCT	calmodulin northern probe
31	calmodulinF-T7	5' - cag aga tgc ata ata cga ctc act ata ggg TGGCAGACAAGTTAACAGAAGAACA	calmodulin northern probe
32	calmodulinR-T7	5' - cag aga tgc ata ata cga ctc act ata ggg TGGCTATCATCATTTTAACAAACTCT	calmodulin northern probe
33	T7-F sequencing primer	5' - TAA TAC GAC TCA CTA TAG GG	sequencing, northern probe
34	MSP1 K1 F (7 bp-tail)	5' - AAA TGA AGA AGA AAT TAC TAC AAA AGG TGC	msp1 genotyping
35	MSP1 K1 R (NED™ -yellow)	5' - GCT TGC ATC AGC TGG AGG GCT TGC ACC AGA	msp1 genotyping
36	MSP1 MAD 20 F (7 bp-tail)	5' - AAA TGA AGG AAC AAG TGG AAC AGC TGT TAC	msp1 genotyping
37	MSP1 MAD 20 R (PET® -red)	5' - ATC TGA AGG ATT TGT ACG TCT TGA ATT ACC	msp1 genotyping
38	MSP1 RO33 F (7 bp-tail)	5' - TAA AGG ATG GAG CAA ATA CTC AAG TTG TTG	msp1 genotyping
39	MSP1 RO33 R (VIC® -green)	5' - CAT CTG AAG GAT TTG CAG CAC CTG GAG ATC	msp1 genotyping
40	MSP2 FC27 F (7 bp-tail)	5' - AAT ACT AAG AGT GTA GGT GCA RAT GCT CCA	msp2 genotyping
41	MSP2 FC27 R (6-FAM™ -blue)	5' - TTT TAT TTG GTG CAT TGC CAG AAC TTG AAC	msp2 genotyping
42	MSP2 IC F (7 bp-tail)	5' - AGA AGT ATG GCA GAA AGT AAK CCT YCT ACT	msp2 genotyping
43	MSP2 IC R (VIC® -green)	5' - GAT TGT AAT TCG GGG GAT TCA GTT TGT TCG	msp2 genotyping
62	PFL2585cloop-L	5' TCC CCC GGG TTA TCC CAA AGA CCA TGT ATT TGC CTC G	PCR, cloning and capillary sequencing
63	PFL2585cloop -R	5' GGC ATG CCA TGG ACT ACA AGG ACG ACG ATG ACA AGT TAT TAA CAG ATG CTG CTG CAA AAC	PCR, cloning and capillary sequencing
64	M13 Forward (-20)	5' - GTA AAA CGA CGG CCA G	sequencing
65	M13 Reverse	5' - CAG GAA ACA GCT ATG AC	sequencing
66	SP6 Forward	5' - GAT TTA GGT GAC ACT ATA G	sequencing

Appendix Method 9.1. *Draft genome assembly and annotation (developed by Thomas Otto).*

From genomic DNA samples of 13 culture-adapted patient isolates, we generated whole genomes draft sequences using the Illumina technology. First we generate noPCR libraries (Kozarewa et al., 2009) with around 500bp fragment length, that were sequenced on the HiSeq result in 100bp reads with a coverage of 60-100x.

After a quick quality control, the reads were *de novo* assembled with velvet (Zerbino and Birney, 2008), k-mer of 61. Next the resulting scaffolds were further scaffolded with SSPACE (Boetzer et al., 2011). The new scaffold were further improved as described in (Swain et al., 2012). In short, the scaffolds were ordered against the *Plasmodium falciparum* reference version 3 (geneDB (Logan-Klumpler et al., 2012)), where the subtelomeric regions were masked with ABACAS (Assefa et al., 2009). Sequencing gaps were closed with image (Tsai et al., 2010).

The genomes were annotated by a combination of transferring the gene models from there reference (RATT (Otto et al., 2011)) and *ab initio* calling by Augustus (Stanke and Morgenstern, 2005). The gene models from the two methods were merged as described in (Otto, 2015).

Newbler configuration:
minimumReadLength 20
overlapSeedStep 12
overlapSeedLength 16
overlapMinSeedCount 1
overlapSeedHitLimit 70
overlapHitPositionLimit 1000000
overlapMinMatchLength 40
overlapMinMatchIdentity 90
overlapMatchIdentScore 2
overlapMatchDiffScore -3
overlapMatchUniqueThresh 12
isogroupThresh 500
isotigThresh 100
isotigContigCountThresh 100
isotigContigLengthThresh 3
aceMode Auto
aceReadMode Default
pairAlignMode None
alignInfoModeAuto
mapMinContigDepth 1
allContigThresh 100
largeContigThresh 500
expectedDepth 0
cDNAMode false
referenceMode Auto
largeGenome false
ripMode false
heterozygoteMode false
assemblerBatchSize 0
numCPU 1
showSingleReadVariations false
nimblegenMappingMode false
backwardCompatibleContigging false
finishMode false
autoTrimming true

Appendix Method 9.1*. Parameters for Newbler assembly of 454 amplicon sequence reads into contigs.*

Appendix Table 9.8. *Raw rif read counts from 454 amplicon sequencing of cDNA and gDNA from laboratory iolates. The data in the table was generated by mapping 454 reads onto homologous genomes and read counts computed using the Artemis application. An average read count was taken when samples were sequenced in duplicates or triplicates. Genes in the table had greater than 5X read coverage.*

Isolate	Gene ID	A/B	ER	LR	ET	MT	LT	S	G
3D7	PF3D7_0421500	A	4	551	1069	1057	491	74	27
3D7	PF3D7_0425900	A	291	587	604	214	289	844	138
3D7	PF3D7_1200500	A	48	108	16	101	700	690	20
3D7	PF3D7_1300400	A	760	440	103	75	72	55	80
3D7	PF3D7_0701100	A	405	464	185	154	149	141	184
3D7	PF3D7_0900400	A	64	392	359	172	162	201	131
3D7	PF3D7_0115600	A	109	379	412	90	111	87	15
3D7	PF3D7_1100300	A	332	223	95	76	55	27	31
3D7	PF3D7_1150300	A	398	195	31	48	44	33	27
3D7	PF3D7_0833400	A	63	79	40	68	164	204	89
3D7	PF3D7_0200200	A	155	137	36	52	45	69	61
3D7	PF3D7_0632700	A	47	86	13	74	134	133	20
3D7	PF3D7_0100200	A	26	91	34	17	131	181	43
3D7	PF3D7_0600500	A	179	109	37	54	54	16	55
3D7	PF3D7_0223400	A	35	64	50	21	154	120	23
3D7	PF3D7_0808900	A	33	142	103	3	63	54	163
3D7	PF3D7_0533000	A	65	146	70	38	50	27	41
3D7	PF3D7_0223100	A	4	110	100	129	34	8	35
3D7	PF3D7_1373400	A	71	55	14	37	95	93	54
3D7	PF3D7_0324800	A	56	58	37	33	92	79	30
3D7	PF3D7_0400300	A	30	68	11	60	64	40	6
3D7	PF3D7_0100400	A	22	45	36	24	90	55	28
3D7	PF3D7_0600300	A	31	28	16	30	51	68	19
3D7	PF3D7_0700200	A	30	65	48	28	9	14	40
3D7	PF3D7_1000500	A	5	25	3	29	53	66	42
3D7	PF3D7_1101100	A	25	47	16	18	38	29	18
3D7	PF3D7_0900600	A	5	23	13	24	47	60	26
3D7	PF3D7_1400200	A	14	33	4	8	73	40	11
3D7	PF3D7_0425600	A	1	43	41	27	37	12	89
3D7	PF3D7_0617700	A	8	35	12	9	49	32	8
3D7	PF3D7_0425700	A	73	23	3	3	18	4	2
3D7	PF3D7_0937700	A	13	21	2	6	36	35	18
3D7	PF3D7_1300200	A	25	20	8	4	29	21	21
3D7	PF3D7_1100500	A	0	36	33	11	18	7	62
3D7	PF3D7_0901300	A	25	21	17	16	5	13	34
3D7	PF3D7_1101300	A	2	25	42	12	6	6	15
3D7	PF3D7_0300200	A	17	22	17	3	19	14	3
3D7	PF3D7_0324400	A	11	17	10	18	28	7	101
3D7	PF3D7_0900200	A	15	14	3	10	29	20	9
3D7	PF3D7_1372700	A	7	11	4	16	16	10	81
3D7	PF3D7_0632200	A	23	13	7	5	4	9	74
3D7	PF3D7_1254700	A	1	21	10	8	13	5	82
3D7	PF3D7_0901500	A	10	16	7	4	3	18	46
3D7	PF3D7_0732900	A	5	6	4	8	23	8	0
3D7	PF3D7_0632400	A	2	22	5	9	10	0	11
3D7	PF3D7_1040500	A	2	23	11	4	4	1	68
3D7	PF3D7_0937500	A	5	5	3	3	16	12	25
3D7	PF3D7_0700300	A	0	10	11	12	6	3	12
3D7	PF3D7_0800400	A	6	10	0	1	14	10	53
3D7	PF3D7_1040300	A	1	8	3	23	5	0	0
3D7	PF3D7_0600700	A	0	6	7	11	7	7	39
3D7	PF3D7_1150000	A	3	9	4	6	9	4	41
3D7	PF3D7_0413200	A	0	5	4	7	8	5	72
3D7	PF3D7_1200200	A	1	7	1	2	9	9	2
3D7	PF3D7_0400700	A	2	7	1	2	4	7	14
3D7	PF3D7_1254200	A	2	5	5	3	2	5	15
3D7	PF3D7_1255100	A	4	7	0	1	2	6	38
3D7	PF3D7_1100400	A	1	6	1	8	4	0	2
3D7	PF3D7_0900700	A	2	7	0	2	6	2	12

Isolate	Gene ID	A/B	ER	LR	ET	MT	LT	S	G
3D7	PF3D7_0400900	A	0	6	3	2	1	1	37
3D7	PF3D7_0732400	A	0	4	3	1	4	0	9
3D7	PF3D7_1150200	A	1	4	0	4	3	0	9
3D7	PF3D7_1479700	A	0	4	1	4	0	3	8
3D7	PF3D7_0100900	A	1	3	2	1	3	1	35
3D7	PF3D7_0400500	A	2	2	1	2	3	1	0
3D7	PF3D7_0300700	A	1	4	0	0	3	2	25
3D7	PF3D7_1041000	A	0	1	0	0	1	8	0
3D7	PF3D7_0712500	A	0	6	0	1	1	1	30
3D7	PF3D7_0732500	A	0	1	2	2	0	4	16
3D7	PF3D7_1253700	A	1	2	0	0	3	2	7
3D7	PF3D7_1400400	A	0	3	0	0	5	0	7
3D7	PF3D7_1300700	A	1	0	2	1	2	2	2
3D7	PF3D7_1480000	A	1	1	0	4	2	0	0
3D7	PF3D7_0300800	A	0	2	1	0	4	0	0
3D7	PF3D7_1254800	A	2	1	2	0	1	0	4
3D7	PF3D7_0712200	A	0	1	0	3	0	2	0
3D7	PF3D7_1301000	A	0	1	0	2	1	1	15
3D7	PF3D7_1254400	A	1	2	0	0	0	2	13
3D7	PF3D7_0833100	A	0	2	0	0	2	0	12
3D7	PF3D7_0200500	A	1	1	1	1	0	0	2
3D7	PF3D7_1040700	A	0	2	0	0	0	1	26
3D7	PF3D7_0401200	A	2	0	0	0	0	0	7
3D7	PF3D7_0222600	A	0	1	0	0	0	0	18
3D7	PF3D7_1000600	B	22	579	354	284	235	162	3
3D7	PF3D7_1300600	B	13	33	10	11	76	130	30
3D7	PF3D7_1372600	B	0	3	0	206	3	0	0
3D7	PF3D7_0900500	B	1	7	5	14	52	132	28
3D7	PF3D7_0701200	B	32	39	14	13	37	17	225
3D7	PF3D7_0100600	B	14	20	9	15	44	31	74
3D7	PF3D7_1479800	B	6	41	50	9	12	5	48
3D7	PF3D7_1000300	B	27	34	7	11	23	11	67
3D7	PF3D7_1400300	B	37	33	5	6	26	3	36
3D7	PF3D7_1149800	B	11	13	32	0	13	6	56
3D7	PF3D7_0201000	B	5	23	6	4	10	9	19
3D7	PF3D7_0101900	B	10	11	2	5	10	13	39
3D7	PF3D7_0324500	B	1	10	9	1	17	11	203
3D7	PF3D7_1400500	B	2	18	3	3	16	1	35
3D7	PF3D7_1255000	B	10	16	1	5	6	2	60
3D7	PF3D7_0900300	B	5	13	4	4	11	3	33
3D7	PF3D7_1479400	B	6	13	4	1	12	3	45
3D7	PF3D7_1040100	B	13	5	1	4	11	3	128
3D7	PF3D7_1400800	B	3	5	3	3	17	4	60
3D7	PF3D7_0500300	B	2	6	2	4	11	8	31
3D7	PF3D7_0732700	B	2	6	9	2	4	4	39
3D7	PF3D7_0300500	B	3	6	1	0	10	7	33
3D7	PF3D7_0632300	B	11	5	1	1	4	3	39
3D7	PF3D7_1040600	B	3	2	4	0	7	8	193
3D7	PF3D7_1041100	B	13	1	1	0	3	3	57
3D7	PF3D7_1040800	B	7	9	2	0	1	0	156
3D7	PF3D7_1040400	B	2	3	7	2	4	1	70
3D7	PF3D7_0833200	B	4	5	1	2	5	1	32
3D7	PF3D7_0901000	B	0	7	2	1	2	3	59
3D7	PF3D7_0631800	B	2	5	0	2	3	3	35
3D7	PF3D7_0832800	B	2	5	0	0	6	2	6
3D7	PF3D7_0532900	B	7	6	0	1	0	0	0
3D7	PF3D7_1040900	B	0	1	1	0	5	6	7
3D7	PF3D7_1254000	B	4	1	1	2	3	0	25
3D7	PF3D7_0222700	B	0	7	1	0	1	0	41
3D7	PF3D7_1101200	B	1	0	3	1	3	1	34
3D7	PF3D7_1254500	B	0	1	0	5	0	0	1
3D7	PF3D7_0114700	B	1	1	0	0	2	0	22
3D7	PF3D7_0222500	B	1	2	0	0	0	1	5

Isolate	Gene ID	A/B	ER	LR	ET	MT	LT	S	G	R-p	T-p
IT	PFIT_0536700	A	3612	4135	2831	1736	2355	1942	157	461	135
IT	PFIT_0900150	A	357	821	1184	2671	3220	1190	34	70	180
IT	PFIT_0835500	A	56	1125	2495	754	755	2408	44	81	128
IT	PFIT_0424300	A	460	1454	153	213	193	155	58	90	81
IT	PFIT_bin02300	A	1000	454	45	22	110	302	54	47	34
IT	PFIT_0537500	A	62	105	23	30	103	324	61	26	5
IT	PFIT_1240200	A	21	48	82	42	60	89	200	1	1
IT	PFIT_bin04400	A	95	81	17	33	58	122	9	8	2
IT	PFIT_1401300	A	122	82	36	10	12	21	124	0	0
IT	PFIT_1401400	A	122	82	36	10	12	21	124	0	0
IT	PFIT_1401500	A	84	51	9	28	29	24	103	0	1
IT	PFIT_0100200	A	165	41	9	10	12	17	1	2	0
IT	PFIT_1400400	A	2	10	31	26	10	21	121	1	1
IT	PFIT_bin08400	A	11	16	32	8	10	13	117	1	1
IT	PFIT_0537200	A	4	24	40	34	16	24	52	0	4
IT	PFIT_bin08800	A	20	26	16	17	19	43	53	1	1
IT	PFIT_0731600	A	15	14	9	6	11	15	77	4	0
IT	PFIT_0537100	A	15	17	11	11	22	28	40	2	3
IT	PFIT_bin05000	A	28	11	5	1	6	5	92	1	0
IT	PFIT_0419700	A	4	6	8	20	14	17	72	1	0
IT	PFIT_1400300	A	0	16	2	2	4	13	91	0	0
IT	PFIT_0411700	A	1	3	1	0	8	8	98	1	0
IT	PFIT_0500500	A	25	17	9	3	6	1	41	5	0
IT	PFIT_bin07100	A	6	3	27	42	10	10	6	0	0
IT	PFIT_bin01100	A	16	15	6	2	7	11	46	0	0
IT	PFIT_bin04900	A	9	18	5	9	6	32	5	9	0
IT	PFIT_0734200	A	0	3	1	0	1	4	82	0	0
IT	PFIT_bin06000	A	2	8	0	3	10	3	62	0	0
IT	PFIT_0200500	A	1	10	6	5	14	13	37	1	0
IT	PFIT_1100200	A	0	4	0	1	5	3	60	0	0
IT	PFIT_bin03700	A	0	0	0	0	0	9	58	0	0
IT	PFIT_0616600	A	6	8	0	3	8	8	28	0	0
IT	PFIT_0731500	A	0	4	14	1	0	6	30	0	0
IT	PFIT_1150500	A	2	4	1	0	2	2	44	0	0
IT	PFIT_bin04600	A	8	2	0	7	9	3	14	5	0
IT	PFIT_1100400	A	0	2	1	0	4	1	39	0	0
IT	PFIT_0300600	A	3	3	1	0	4	2	33	0	0
IT	PFIT_0800400	A	3	2	3	0	2	2	33	0	0
IT	PFIT_0733300	A	3	0	0	1	6	1	31	0	0
IT	PFIT_bin03500	A	2	5	2	9	8	2	13	0	0
IT	PFIT_0800700	A	5	4	4	4	10	9	1	2	0
IT	PFIT_0536500	A	8	4	2	8	2	5	5	0	0
IT	PFIT_bin02800	A	3	1	4	0	0	4	20	0	0
IT	PFIT_0424000	A	2	5	0	0	1	1	21	0	0
IT	PFIT_0200200	A	2	1	0	2	3	1	20	0	0
IT	PFIT_0801500	A	0	6	3	0	1	1	18	0	0
IT	PFIT_1150200	A	1	6	2	1	2	1	13	0	0
IT	PFIT_0400100	A	0	1	0	0	0	0	22	1	0
IT	PFIT_bin06700	A	0	2	5	3	1	3	9	0	0
IT	PFIT_bin10200	A	2	4	5	0	8	0	0	0	0
IT	PFIT_0300700	A	3	2	1	1	1	4	5	0	0
IT	PFIT_bin05900	A	0	1	0	0	4	1	6	0	0
IT	PFIT_0500400	A	0	0	1	2	3	0	4	0	0
IT	PFIT_bin02900	A	2	1	3	0	0	1	2	0	0
IT	PFIT_0411800	A	0	0	0	0	1	2	5	0	0
IT	PFIT_bin00500	A	1	0	1	0	6	0	0	0	0
IT	PFIT_0901100	A	2	3	0	0	1	0	0	0	0
IT	PFIT_0700400	A	0	0	0	1	0	1	3	0	0
IT	PFIT_0901700	A	0	1	0	0	0	0	4	0	0
IT	PFIT_bin03400	A	0	0	2	2	0	0	1	0	0
IT	PFIT_bin06800	A	2	3	0	0	0	0	0	0	0
IT	PFIT_bin07900	A	1	0	1	0	3	0	0	0	0
IT	PFIT_bin08000	A	1	0	1	0	3	0	0	0	0
IT	PFIT_bin08100	A	1	0	1	0	3	0	0	0	0
IT	PFIT_0537000	A|ps	12	20	10	11	20	34	36	2	4

Isolate	Gene ID	A/B	ER	LR	ET	MT	LT	S	G	R-p	T-p
IT	PFIT_0733200	B	7	109	101	256	572	82	1	0	1
IT	PFIT_1150600	B	10	4	3	5	23	106	64	22	0
IT	PFIT_0100400	B	13	24	0	6	13	3	102	37	0
IT	PFIT_bin04800	B	25	30	3	9	17	17	88	7	0
IT	PFIT_0424100	B	19	22	4	17	7	8	108	5	1
IT	PFIT_0700600	B	8	11	1	14	20	14	108	0	1
IT	PFIT_0300400	B	4	8	11	11	6	14	98	0	1
IT	PFIT_0300500	B	4	8	11	11	6	14	98	0	1
IT	PFIT_bin03600	B	7	15	2	13	13	9	65	7	0
IT	PFIT_0800300	B	5	15	12	14	5	7	67	0	0
IT	PFIT_0901200	B	6	9	2	6	5	1	95	1	0
IT	PFIT_0733500	B	1	5	8	7	4	1	93	0	0
IT	PFIT_0901300	B	6	3	0	4	2	1	95	0	0
IT	PFIT_bin03900	B	0	1	1	3	1	0	89	0	1
IT	PFIT_bin10000	B	21	1	0	3	3	6	61	0	0
IT	PFIT_bin01200	B	0	7	3	2	2	4	68	0	1
IT	PFIT_0537400	B	11	0	1	0	3	4	25	2	0
IT	PFIT_0536600	B	4	3	0	3	2	6	13	1	4
IT	PFIT_0113500	B	1	3	0	2	0	0	28	0	0
IT	PFIT_bin01700	B	0	1	0	0	0	0	21	0	0
IT	PFIT_bin00800	B	0	0	2	2	0	0	8	0	0
IT	PFIT_0200300	B	0	3	0	1	1	1	5	0	0
IT	PFIT_0733400	B	1	0	0	2	1	5	0	0	0
IT	PFIT_bin10400	B	0	0	1	0	0	1	7	0	0

Isolate	Gene ID	A/B	ER	LR	ET	MT	LT	S	G
HB3	PfHB3_BroadWash530300	A	1227	686	1235	808	697	958	165
HB3	PfHB3_BroadWash520600	A	311	294	240	290	166	239	163
HB3	PfHB3_BroadWash536100	A	229	94	158	158	273	223	177
HB3	PfHB3_BroadWash524200	A	170	93	134	189	136	226	57
HB3	PfHB3_BroadWash524700	A	247	123	160	167	136	157	13
HB3	PfHB3_BroadWash534900	A	58	65	141	158	61	159	33
HB3	PfHB3_BroadWash162300	A	94	76	166	127	18	105	81
HB3	PfHB3_BroadWash523800	A	69	82	81	134	68	144	80
HB3	PfHB3_BroadWash521900	A	14	46	89	76	12	168	49
HB3	PfHB3_BroadWash520100	A	41	68	75	90	35	104	39
HB3	PfHB3_BroadWash074000	A	8	49	63	55	121	104	49
HB3	PfHB3_BroadWash075100	A	0	54	97	58	20	89	104
HB3	PfHB3_BroadWash172400	A	9	38	56	64	17	73	114
HB3	PfHB3_BroadWash520900	A	0	45	60	72	20	85	85
HB3	PfHB3_BroadWash539200	A	20	76	57	42	32	72	18
HB3	PfHB3_BroadWash526000	A	5	70	60	36	5	96	34
HB3	PfHB3_BroadWash161400	A	6	57	93	60	8	67	12
HB3	PfHB3_BroadWash511900	A	53	22	28	36	64	43	47
HB3	PfHB3_BroadWash363600	A	8	18	11	29	28	24	159
HB3	PfHB3_BroadWash520700	A	62	29	38	40	21	29	42
HB3	PfHB3_BroadWash516900	A	25	30	48	39	1	52	61
HB3	PfHB3_BroadWash525700	A	0	6	15	9	2	17	165
HB3	PfHB3_BroadWash509300	A	29	37	36	35	13	31	11
HB3	PfHB3_BroadWash530100	A	3	21	26	41	9	55	23
HB3	PfHB3_BroadWash578000	A	34	7	13	19	13	39	24
HB3	PfHB3_BroadWash528900	A	8	3	6	11	17	46	56
HB3	PfHB3_BroadWash522500	A	8	20	23	20	18	31	7
HB3	PfHB3_BroadWash531500	A	2	9	5	6	9	19	62
HB3	PfHB3_BroadWash515900	A	0	13	19	14	3	22	27
HB3	PfHB3_BroadWash531000	A	2	9	7	11	1	3	65
HB3	PfHB3_BroadWash528800	A	1	7	12	21	4	11	25
HB3	PfHB3_BroadWash193800	A	0	2	1	0	8	4	65
HB3	PfHB3_BroadWash512400	A	5	14	10	12	4	19	16
HB3	PfHB3_BroadWash512200	A	0	7	20	11	2	11	22
HB3	PfHB3_BroadWash556100	A	0	8	13	5	0	7	31
HB3	PfHB3_BroadWash135100	A	0	1	2	0	2	18	36
HB3	PfHB3_BroadWash526100	A	0	6	6	8	6	8	24
HB3	PfHB3_BroadWash545000	A	1	13	5	4	0	14	18
HB3	PfHB3_BroadWash544900	A	6	1	1	7	6	5	21
HB3	PfHB3_BroadWash071100	A	2	3	3	5	2	3	21
HB3	PfHB3_BroadWash538700	A	1	2	0	3	0	1	32
HB3	PfHB3_BroadWash531200	A	7	0	2	2	4	5	18
HB3	PfHB3_BroadWash534600	A	0	1	3	2	1	5	24
HB3	PfHB3_BroadWash509800	A	0	1	2	0	0	3	24
HB3	PfHB3_BroadWash539400	A	0	1	0	2	0	4	18
HB3	PfHB3_BroadWash531300	A	5	0	2	0	8	4	4
HB3	PfHB3_BroadWash011500	A	4	1	1	0	12	1	2
HB3	PfHB3_BroadWash172300	A	0	1	0	1	1	4	14
HB3	PfHB3_BroadWash509500	A	1	0	2	1	4	9	2
HB3	PfHB3_BroadWash541200	A	0	3	1	4	1	2	8
HB3	PfHB3_BroadWash538400	A	5	0	0	1	1	6	5
HB3	PfHB3_BroadWash541500	A	1	2	3	5	0	2	4
HB3	PfHB3_BroadWash578300	A	7	0	1	1	4	3	0
HB3	PfHB3_BroadWash519900	A	0	0	0	1	1	6	7
HB3	PfHB3_BroadWash161900	A	1	0	0	0	5	0	6
HB3	PfHB3_BroadWash527700	A	2	0	1	0	5	2	0
HB3	PfHB3_BroadWash526400	A	1	1	0	1	0	2	4
HB3	PfHB3_BroadWash526800	A	0	0	0	2	1	2	4
HB3	PfHB3_BroadWash523500	A	2	0	0	0	2	0	3
HB3	PfHB3_BroadWash542500	A	0	2	3	0	1	1	0
HB3	PfHB3_BroadWash192400	A	0	0	0	0	0	0	6
HB3	PfHB3_BroadWash531100	A	0	1	1	1	0	0	3
HB3	PfHB3_BroadWash550900	A	0	0	1	0	5	0	0
HB3	PfHB3_BroadWash162100	A	0	1	0	2	1	1	0

Isolate	Gene ID	A/B	ER	LR	ET	MT	LT	S	G
HB3	PfHB3_BroadWash530200	B	65	340	350	298	340	308	75
HB3	PfHB3_BroadWash551000	B	55	168	185	194	36	176	131
HB3	PfHB3_BroadWash161200	B	20	26	30	31	99	83	385
HB3	PfHB3_BroadWash526600	B	11	27	37	37	12	57	462
HB3	PfHB3_BroadWash262300	B	11	58	83	83	29	152	46
HB3	PfHB3_BroadWash509700	B	10	16	28	26	3	34	133
HB3	PfHB3_BroadWash526200	B	2	31	29	42	7	47	84
HB3	PfHB3_BroadWash539500	B	2	21	31	37	8	49	85
HB3	PfHB3_BroadWash529000	B	26	4	5	13	49	14	117
HB3	PfHB3_BroadWash516500	B	0	3	8	16	4	6	185
HB3	PfHB3_BroadWash521500	B	1	25	41	39	13	14	62
HB3	PfHB3_BroadWash509400	B	4	4	5	16	27	8	74
HB3	PfHB3_BroadWash531700	B	2	1	0	2	3	3	105
HB3	PfHB3_BroadWash539300	B	15	0	4	2	2	4	21
HB3	PfHB3_BroadWash528700	B	6	0	1	2	5	2	12
HB3	PfHB3_BroadWash541300	B	1	0	0	1	2	3	20
HB3	PfHB3_BroadWash537700	B	0	1	1	0	2	1	18
HB3	PfHB3_BroadWash519800	B	1	2	4	3	4	6	0
HB3	PfHB3_BroadWash512000	B	0	0	0	3	1	1	4
HB3	PfHB3_BroadWash532400	B	1	0	1	1	1	0	3
HB3	PfHB3_BroadWash536000	B	1	0	1	1	3	0	1
HB3	PfHB3_BroadWash520000	B	0	0	0	0	2	4	0
HB3	PfHB3_BroadWash526300	B	0	0	0	0	0	3	2

Appendix Table 9.9. *Raw stevor read counts from 454 amplicon sequencing of cDNA and gDNA from laboratory isolates. The data in the table was generated by mapping 454 reads onto homologous genomes and computing read counts using Artemis. An average read count was taken when samples were sequenced in duplicates or triplicates. Genes in the table had greater than 5X read coverage.*

Isolate	Gene ID	ER	LR	ET	MT	LT	S	G
3D7	PF3D7_1040200	106	1424	836	1518	772	648	26
3D7	PF3D7_0617600	453	765	377	506	697	682	71
3D7	PF3D7_1479500	106	426	216	161	208	284	134
3D7	PF3D7_0401500	292	251	28	18	94	345	45
3D7	PF3D7_1149900	133	178	62	140	174	169	82
3D7	PF3D7_0901600	61	37	11	61	209	445	48
3D7	PF3D7_0732000	202	183	24	33	76	190	158
3D7	PF3D7_0631900	30	145	142	204	117	105	68
3D7	PF3D7_1300900	63	313	126	37	20	83	113
3D7	PF3D7_0324600	64	212	99	37	65	136	135
3D7	PF3D7_0101800	68	57	16	17	24	192	198
3D7	PF3D7_0400800	97	92	28	13	51	108	146
3D7	PF3D7_1400700	128	102	30	25	41	71	94
3D7	PF3D7_1479900	88	40	24	80	40	58	43
3D7	PF3D7_1254600	32	70	46	55	27	54	70
3D7	PF3D7_0900900	58	36	7	21	35	68	113
3D7	PF3D7_0300400	66	41	11	3	29	24	116
3D7	PF3D7_0200900	40	42	1	3	26	62	113
3D7	PF3D7_0700400	37	46	4	9	16	85	81
3D7	PF3D7_0115400	18	79	36	21	21	30	58
3D7	PF3D7_0201300	29	42	9	7	22	87	20
3D7	PF3D7_1254100	36	31	15	19	24	40	40
3D7	PF3D7_0832400	44	26	22	11	21	42	33
3D7	PF3D7_0832900	21	23	33	36	19	38	29
3D7	PF3D7_0222800	21	21	3	0	18	15	119
3D7	PF3D7_0114600	25	18	2	6	18	24	47
3D7	PF3D7_0200400	21	12	1	1	8	8	15
3D7	PF3D7_1254300	1	2	2	4	1	0	14
IT	PFIT_bin03800	139	180	1727	3908	2132	2527	188
IT	PFIT_0400200	459	1049	450	383	269	518	206
IT	PFIT_0700100	372	189	487	221	205	415	331
IT	PFIT_0100300	135	103	111	392	354	383	375
IT	PFIT_bin00600	325	207	277	131	123	192	515
IT	PFIT_0900800	156	55	20	13	97	118	564
IT	PFIT_0900900	156	55	20	13	97	118	564
IT	PFIT_0835400	142	63	121	153	166	191	115
IT	PFIT_1300300	185	125	16	27	86	77	384
IT	PFIT_0800600	30	59	15	20	99	411	180
IT	PFIT_0700500	110	71	12	20	67	102	370
IT	PFIT_0733800	33	15	100	114	86	171	200
IT	PFIT_1100300	159	79	41	42	44	65	212
IT	PFIT_bin05300	42	32	14	30	26	31	363
IT	PFIT_1401700	91	51	10	9	43	29	249
IT	PFIT_0731100	79	29	14	7	39	67	195
IT	PFIT_0113400	53	40	11	5	14	49	228
IT	PFIT_bin09500	31	34	5	4	21	12	190
IT	PFIT_0734100	8	3	10	11	15	14	183
IT	PFIT_bin03100	58	26	29	12	18	15	73
IT	PFIT_bin03200	58	26	29	12	18	15	73
IT	PFIT_bin03300	58	26	29	12	18	15	73
IT	PFIT_bin01600	41	5	4	4	10	5	103
IT	PFIT_0300100	13	2	3	16	22	36	50
IT	PFIT_0200600	2	2	7	2	2	5	7
IT	PFIT_0100600	0	0	0	0	0	0	5

Isolate	Gene ID	ER	LR	ET	MT	LT	S	G
HB3	PfHB3_BroadWash161300	117	460	506	627	4182	869	275
HB3	PfHB3_BroadWash532500	99	137	115	191	4861	373	227
HB3	PfHB3_BroadWash521400	15	699	648	456	1743	569	208
HB3	PfHB3_BroadWash538600	46	57	64	71	3167	157	159
HB3	PfHB3_BroadWash544800	54	279	355	381	1583	551	70
HB3	PfHB3_BroadWash534700	15	297	252	232	1263	168	122
HB3	PfHB3_BroadWash528600	68	139	238	258	648	300	51
HB3	PfHB3_BroadWash549600	38	80	134	137	559	122	176
HB3	PfHB3_BroadWash530000	52	14	14	48	774	65	9
HB3	PfHB3_BroadWash162000	71	2	2	10	797	30	3
HB3	PfHB3_BroadWash531600	9	38	43	91	433	119	131
HB3	PfHB3_BroadWash509900	9	15	37	20	383	68	84
HB3	PfHB3_BroadWash509600	10	34	27	26	318	60	54
HB3	PfHB3_BroadWash537600	7	30	24	23	294	37	73
HB3	PfHB3_BroadWash516600	2	9	13	13	165	8	10
HB3	PfHB3_BroadWash523400	0	1	3	2	44	2	0
HB3	PfHB3_BroadWash529200	0	0	0	0	8	1	0
HB3	PfHB3_BroadWash541100	0	0	0	0	9	0	0

Appendix Table 9.10. *Raw rif read counts from 454 amplicon sequencing of cDNA and gDNA from patient iolates that were put under differential selection pressure. AC (acute), CU (culture-adapted) HB (HBEC-slelected), R+ (rosetting clone), R- (non-rosetting fraction). The data in the table was generated by mapping 454 reads onto homologous genomes and computing read counts using Artemis. An average read count was taken when samples were sequenced in duplicates or triplicates. Genes in the table had greater than 5X read coverage.*

Isolate	Gene ID	A/B	AC_R	AC T	AC G	CU_R	CU_LRET	CU_T	CU_S	CU_G	
9215	PF8443_8_4_U509100	A	87	68	40	197	230	85	85	20	811
9215	PF8443_8_4_U518700	A	99	117	332	10	8	10	7	29	610
9215	PF8443_8_4_U517400	A	63	103	113	44	22	135	58	52	590
9215	PF8443_8_4_U511800	A	28	125	91	12	19	41	71	34	420
9215	PF8443_8_4_U071400	A	9	81	140	6	56	24	38	65	417
9215	PF8443_8_4_U519000	A	126	84	141	11	10	1	4	7	383
9215	PF8443_8_4_U515700	A	42	152	79	11	9	31	30	18	371
9215	PF8443_8_4_U515000	A	44	75	31	21	65	20	82	10	345
9215	PF8443_8_4_U552000	A	17	53	159	5	3	1	7	69	313
9215	PF8443_8_4_U504400	A	118	104	17	14	12	4	7	11	286
9215	PF8443_8_4_U546100	A	23	20	116	19	15	15	10	59	277
9215	PF8443_8_4_U520100	A	35	37	104	14	18	9	11	37	264
9215	PF8443_8_4_U530100	A	33	76	48	17	27	16	22	22	260
9215	PF8443_8_4_U520300	A	9	79	85	9	11	20	14	24	250
9215	PF8443_8_4_U195200	A	10	16	94	9	14	17	23	43	224
9215	PF8443_8_4_U067000	A	7	40	56	14	29	15	12	19	191
9215	PF8443_8_4_U549800	A	15	41	35	21	24	17	12	19	183
9215	PF8443_8_4_U503200	A	25	4	85	10	8	12	6	33	181
9215	PF8443_8_4_U134400	A	13	58	38	8	10	8	24	16	174
9215	PF8443_8_4_U546800	A	1	0	8	8	14	90	50	3	173
9215	PF8443_8_4_U560200	A	13	60	15	11	14	13	16	23	165
9215	PF8443_8_4_U559500	A	5	1	6	125	15	2	2	3	158
9215	PF8443_8_4_U557300	A	46	22	24	6	12	10	14	17	149
9215	PF8443_8_4_U560400	A	17	46	30	8	11	9	12	14	146
9215	PF8443_8_4_U519200	A	12	13	83	1	3	2	3	28	144
9215	PF8443_8_4_U511300	A	7	22	66	4	7	3	11	25	143
9215	PF8443_8_4_U519100	A	11	10	88	2	1	2	4	26	142
9215	PF8443_8_4_U502800	A	10	11	59	10	10	4	6	25	134
9215	PF8443_8_4_U515800	A	45	32	49	2	1	1	2	1	132
9215	PF8443_8_4_U515300	A	12	5	60	4	2	3	14	20	119
9215	PF8443_8_4_U517600	A	12	29	40	3	6	11	8	11	119
9215	PF8443_8_4_U161400	A	27	25	29	6	5	3	2	7	103
9215	PF8443_8_4_U561400	A	12	17	33	4	2	3	5	14	89
9215	PF8443_8_4_U543000	A	1	0	20	11	30	11	7	8	86
9215	PF8443_8_4_U519300	A	21	16	20	6	5	4	8	6	85
9215	PF8443_8_4_U059500	A	1	0	64	1	0	2	2	14	84
9215	PF8443_8_4_U518500	A	24	7	42	1	4	0	1	4	82
9215	PF8443_8_4_U161600	A	3	0	45	1	4	2	2	26	81
9215	PF8443_8_4_U555200	A	4	21	14	4	16	9	7	7	81
9215	PF8443_8_4_U561300	A	7	4	23	7	11	8	14	7	79
9215	PF8443_8_4_U161700	A	8	7	27	20	4	3	3	8	79
9215	PF8443_8_4_U560300	A	9	24	16	7	3	7	4	6	76
9215	PF8443_8_4_U504500	A	12	12	22	4	4	4	9	11	76
9215	PF8443_8_4_U534200	A	6	11	18	2	11	11	8	7	73
9215	PF8443_8_4_U509500	A	4	42	10	3	2	4	3	4	71
9215	PF8443_8_4_U510300	A	11	0	24	10	5	9	4	10	71
9215	PF8443_8_4_U503100	A	1	6	38	2	3	2	2	14	65
9215	PF8443_8_4_U518100	A	5	15	5	2	9	14	9	5	62
9215	PF8443_8_4_U516000	A	7	3	29	5	1	4	2	6	57
9215	PF8443_8_4_U560100	A	18	19	13	1	2	0	2	2	56
9215	PF8443_8_4_U548600	A	6	5	30	0	3	2	0	6	51
9215	PF8443_8_4_U502900	A	1	3	30	2	1	2	2	11	51
9215	PF8443_8_4_U517800	A	5	8	19	3	1	3	2	10	49
9215	PF8443_8_4_U070100	A	9	1	16	0	6	4	3	10	49
9215	PF8443_8_4_U511400	A	4	5	20	2	2	2	0	14	48
9215	PF8443_8_4_U512300	A	4	10	11	3	6	4	4	5	46
9215	PF8443_8_4_U530000	A	8	3	11	1	8	5	3	8	46
9215	PF8443_8_4_U546200	A	3	0	11	4	0	3	4	12	36
9215	PF8443_8_4_U548500	A	1	1	8	0	1	4	2	3	19
9215	PF8443_8_4_U504800	A	0	2	4	2	3	1	1	2	14
9215	PF8443_8_4_U557200	A	5	1	3	0	2	1	1	1	14
9215	PF8443_8_4_U162100	A	3	0	5	1	2	0	0	2	12
9215	PF8443_8_4_U549100	A	1	0	6	0	1	2	0	3	12
9215	PF8443_8_4_U161500	A	0	0	9	0	0	1	1	0	10
9215	PF8443_8_4_U530400	A	2	0	2	2	1	1	2	1	9

Isolate	Gene ID	A/B	AC_R	AC T	AC G	CU_R	CU_LRET	CU_T	CU_S	CU_G	
9215	PF8443_8_4_U534100	A	0	0	3	1	1	2	1	2	8
9215	PF8443_8_4_U518200	A	1	0	5	0	0	1	1	0	7
9215	PF8443_8_4_U518000	A	1	1	3	1	1	0	0	1	7
9215	PF8443_8_4_U503000	A	0	0	2	1	0	1	0	3	7
9215	PF8443_8_4_U549600	A	1	1	1	1	1	0	0	0	5
9215	PF8443_8_4_U560800	A	0	1	1	0	1	1	0	1	4
9215	PF8443_8_4_U069900	A	3	0	1	0	1	0	0	0	4
9215	PF8443_8_4_U532500	A	0	0	1	3	0	0	0	0	4
9215	PF8443_8_4_U539500	A	177	282	103	573	344	249	207	27	1962
9215	PF8443_8_4_U521600	A	96	362	117	45	47	162	356	31	1215
9215	PF8443_8_4_U522000	A	41	25	59	157	204	104	122	20	732
9215	PF8443_8_4_U522700	A	107	108	135	46	54	36	63	26	573
9215	PF8443_8_4_U522200	A	91	107	150	3	2	2	3	5	361
9215	PF8443_8_4_U521100	A	77	42	33	5	43	15	14	14	241
9215	PF8443_8_4_U521300	A	16	73	79	3	3	10	10	27	220
9215	PF8443_8_4_U520800	A	78	15	12	38	29	21	11	3	206
9215	PF8443_8_4_U521800	A	48	0	20	12	29	9	12	10	139
9215	PF8443_8_4_U521200	A	32	23	65	1	1	3	1	4	129
9215	PF8443_8_4_U522300	A	3	11	12	4	17	18	27	5	96
9215	PF8443_8_4_U520600	A	7	44	20	2	1	2	4	12	90
9215	PF8443_8_4_U541400	A	7	4	25	2	5	1	2	8	53
9215	PF8443_8_4_U550500	A	1	22	10	3	0	0	2	13	51
9215	PF8443_8_4_U534400	A	6	4	19	1	3	2	3	13	49
9215	PF8443_8_4_U526100	A	5	4	4	9	13	5	3	1	42
9215	PF8443_8_4_U523700	A	1	10	11	4	2	1	2	6	37
9215	PF8443_8_4_U541300	A	6	3	9	3	4	4	3	5	36
9215	PF8443_8_4_U540500	A	1	4	9	1	2	0	0	1	17
9215	PF8443_8_4_U515100	A	0	0	3	0	1	0	0	1	5
9215	PF8443_8_4_U512100	A ps	57	53	59	263	273	175	189	34	1102
9215	PF8443_8_4_U525300	A trunc	17	9	31	7	8	10	7	6	93
9215	PF8443_8_4_U526400	A trunc	16	19	3	2	3	2	1	3	49
9215	PF8443_8_4_U504700	A trunc	4	2	13	2	1	4	1	4	30
9215	PF8443_8_4_U525900	A trunc	1	0	2	0	0	1	0	2	6
9215	PF8443_8_4_U538200	B	218	117	130	153	457	424	168	59	1725
9215	PF8443_8_4_U505500	B	101	78	142	42	46	31	47	34	520
9215	PF8443_8_4_U537200	B	52	132	93	48	50	25	32	38	468
9215	PF8443_8_4_U530500	B	71	72	117	47	33	19	22	61	440
9215	PF8443_8_4_U530200	B	26	83	75	32	43	61	37	34	389
9215	PF8443_8_4_U520500	B	31	40	84	34	51	27	28	55	349
9215	PF8443_8_4_U549700	B	31	6	86	66	61	30	28	29	337
9215	PF8443_8_4_U506700	B	25	61	120	18	14	26	19	51	333
9215	PF8443_8_4_U510600	B	53	39	118	15	11	28	17	44	324
9215	PF8443_8_4_U509200	B	44	38	50	26	14	15	22	57	265
9215	PF8443_8_4_U161800	B	25	27	93	20	18	18	17	42	259
9215	PF8443_8_4_U541500	B	7	40	105	8	7	11	10	41	228
9215	PF8443_8_4_U505600	B	12	104	34	8	12	11	20	22	222
9215	PF8443_8_4_U507500	B	13	19	93	2	9	9	9	42	196
9215	PF8443_8_4_U552100	B	38	24	79	4	5	8	5	24	186
9215	PF8443_8_4_U548400	B	23	12	48	20	16	17	6	31	171
9215	PF8443_8_4_U510400	B	11	13	53	2	4	1	3	27	112
9215	PF8443_8_4_U508000	B	8	23	39	3	6	10	8	9	104
9215	PF8443_8_4_U559200	B	8	9	40	3	3	6	2	27	96
9215	PF8443_8_4_U559800	B	10	15	19	6	6	9	10	11	85
9215	PF8443_8_4_U512200	B	8	4	13	6	3	4	7	12	56
9215	PF8443_8_4_U507800	B	3	6	15	4	2	2	2	17	50
9215	PF8443_8_4_U162000	B	5	1	4	5	3	1	6	7	30
9215	PF8443_8_4_U551800	B	2	1	6	4	1	1	3	8	25
9215	PF8443_8_4_U545500	B	2	4	13	1	0	0	0	4	24
9215	PF8443_8_4_U161900	B	1	1	8	1	0	0	1	4	14
9215	PF8443_8_4_U546500	B	1	1	5	0	0	1	1	4	12
9215	PF8443_8_4_U555100	B	2	1	4	1	1	1	0	2	12
9215	PF8443_8_4_U557600	B	0	0	6	1	0	0	0	1	8
9215	PF8443_8_4_U505900	B	1	1	4	1	1	0	0	0	7
9215	PF8443_8_4_U265100	B	0	1	2	1	2	0	0	1	6
9215	PF8443_8_4_U552400	B	2	1	0	1	0	1	0	1	5
9215	PF8443_8_4_U547000	B	0	0	1	1	1	1	1	1	4

Isolate	Gene ID	A/B	AC_R	AC_G	CU_R2	CU_P1	CU_T1	CU_S1	CU_G	HB_R1	HB_P1	HB_T1	HB_S1	HB_G
9106	PF8443_8_7_U503900	A	8	87	50	1516	2112	1176	25	2	4	2	23	110
9106	PF8443_8_7_U525700	A	182	95	203	405	188	178	46	501	319	1	455	283
9106	PF8443_8_7_U484300	A	93	48	62	388	470	276	21	457	197	9	334	111
9106	PF8443_8_7_U513900	A	115	59	98	374	395	251	32	79	59	2	33	154
9106	PF8443_8_7_U439800	A	193	67	97	11	10	14	29	595	414	0	46	89
9106	PF8443_8_7_U541800	A	66	92	15	99	106	92	46	268	269	4	208	224
9106	PF8443_8_7_U436600	A	8	58	5	324	443	239	25	5	23	4	57	106
9106	PF8443_8_7_U537000	A	25	98	6	232	230	150	62	7	13	0	43	225
9106	PF8443_8_7_U515900	A	86	100	37	126	140	90	58	67	63	16	76	152
9106	PF8443_8_7_U518900	A	76	92	35	55	103	78	66	136	78	1	73	207
9106	PF8443_8_7_U525800	A	21	33	16	145	166	103	19	70	46	6	272	76
9106	PF8443_8_7_U564900	A	35	47	72	14	2	6	18	187	495	0	14	77
9106	PF8443_8_7_U527200	A	111	45	29	34	24	18	14	163	269	2	87	53
9106	PF8443_8_7_U443600	A	11	41	54	221	98	66	27	4	9	5	91	84
9106	PF8443_8_7_U513600	A	11	145	1	25	43	29	64	2	22	2	90	277
9106	PF8443_8_7_U447300	A	29	116	4	81	63	48	63	7	12	5	25	248
9106	PF8443_8_7_U524900	A	70	120	3	33	34	19	68	16	14	1	52	258
9106	PF8443_8_7_U560800	A	24	115	7	38	47	23	55	17	18	2	24	215
9106	PF8443_8_7_U536000	A	9	29	2	17	20	23	17	104	73	9	196	65
9106	PF8443_8_7_U448300	A	106	70	54	36	33	34	30	4	46	0	22	108
9106	PF8443_8_7_U458100	A	7	62	1	3	4	5	57	1	2	1	116	199
9106	PF8443_8_7_U525300	A	36	41	21	48	24	37	23	20	17	0	72	116
9106	PF8443_8_7_U503400	A	31	76	9	9	4	4	17	71	74	0	17	142
9106	PF8443_8_7_U462400	A	4	122	0	3	13	10	42	1	0	0	6	251
9106	PF8443_8_7_U051200	A	18	115	0	29	27	19	40	7	6	0	48	137
9106	PF8443_8_7_U554100	A	5	120	0	1	0	0	44	16	1	0	0	236
9106	PF8443_8_7_U490200	A	10	13	100	75	42	57	16	8	28	0	7	40
9106	PF8443_8_7_U462300	A	4	74	5	21	32	31	39	7	18	2	3	149
9106	PF8443_8_7_U448200	A	59	53	12	27	10	3	25	6	28	0	3	116
9106	PF8443_8_7_U515700	A	16	77	8	13	6	12	24	15	27	0	35	103
9106	PF8443_8_7_U112500	A	22	69	1	28	20	21	24	3	2	0	63	79
9106	PF8443_8_7_U100500	A	6	24	2	45	57	67	10	1	40	1	49	25
9106	PF8443_8_7_U523500	A	7	57	11	27	21	15	26	4	55	0	22	80
9106	PF8443_8_7_U484900	A	5	56	7	26	14	18	27	2	17	1	36	112
9106	PF8443_8_7_U460800	A	27	72	0	15	11	16	43	2	0	0	3	127
9106	PF8443_8_7_U435600	A	5	62	8	45	43	30	20	0	1	1	23	74
9106	PF8443_8_7_U520200	A	68	22	19	25	14	11	18	15	14	0	15	62
9106	PF8443_8_7_U521400	A	14	30	1	20	15	10	41	5	15	1	11	104
9106	PF8443_8_7_U530500	A	47	71	15	37	16	21	19	9	5	0	1	23
9106	PF8443_8_7_U558000	A	16	24	4	15	21	14	9	23	53	1	19	42
9106	PF8443_8_7_U525200	A	23	28	5	10	12	13	18	35	8	0	6	71
9106	PF8443_8_7_U441300	A	22	43	0	6	0	2	21	21	26	0	21	67
9106	PF8443_8_7_U554200	A	18	49	0	9	11	22	17	4	1	1	10	81
9106	PF8443_8_7_U541300	A	61	43	6	2	2	1	17	12	7	0	13	56
9106	PF8443_8_7_U439900	A	16	50	4	6	10	11	25	1	1	0	17	78
9106	PF8443_8_7_U561800	A	31	52	9	8	15	15	26	19	7	1	10	21
9106	PF8443_8_7_U438600	A	4	36	0	23	35	16	8	0	1	0	19	51
9106	PF8443_8_7_U551000	A	14	43	0	2	4	4	19	11	3	0	13	71
9106	PF8443_8_7_U446000	A	6	36	1	6	14	12	23	0	2	0	3	77
9106	PF8443_8_7_U520300	A	3	37	0	16	20	16	17	0	0	0	5	67
9106	PF8443_8_7_U437700	A	16	41	3	2	6	2	18	13	8	0	3	67
9106	PF8443_8_7_U482200	A	11	26	0	11	23	5	9	24	18	1	9	41
9106	PF8443_8_7_U560900	A	5	32	4	14	22	26	8	0	0	0	6	35
9106	PF8443_8_7_U514500	A	9	15	2	7	7	6	8	18	34	0	17	27
9106	PF8443_8_7_U515300	A	2	38	1	5	4	3	20	2	0	0	5	50
9106	PF8443_8_7_U441200	A	8	28	0	6	2	6	5	6	5	0	41	20
9106	PF8443_8_7_U439500	A	3	23	1	5	0	4	15	2	3	0	9	45
9106	PF8443_8_7_U482500	A	0	15	0	2	13	4	10	0	3	0	13	45
9106	PF8443_8_7_U529100	A	8	25	0	2	4	4	13	2	2	0	2	35
9106	PF8443_8_7_U472000	A	7	16	4	0	1	0	11	4	3	0	5	43
9106	PF8443_8_7_U488500	A	29	13	5	3	4	0	4	9	2	0	10	12

Isolate	Gene ID	A/B	AC_R	AC_G	CU_R2	CU_P1	CU_T1	CU_S1	CU_G	HB_R1	HB_P1	HB_T1	HB_S1	HB_G
9106	PF8443_8_7_U557200	A	0	15	0	1	0	1	1	14	8	1	39	9
9106	PF8443_8_7_U443500	A	7	12	0	9	10	8	6	3	5	0	7	17
9106	PF8443_8_7_U473600	A	5	10	1	2	3	3	7	12	15	0	3	19
9106	PF8443_8_7_U559400	A	1	9	0	11	7	4	5	3	13	0	11	11
9106	PF8443_8_7_U560600	A	4	10	7	1	1	3	6	0	4	1	12	13
9106	PF8443_8_7_U436200	A	2	11	0	2	1	2	9	1	0	0	4	24
9106	PF8443_8_7_U460700	A	19	7	2	0	0	1	2	5	2	0	3	6
9106	PF8443_8_7_U502900	A	2	8	0	4	3	3	2	0	0	0	1	8
9106	PF8443_8_7_U471700	A	1	5	0	0	1	2	5	0	3	0	1	10
9106	PF8443_8_7_U518400	A	1	4	0	5	4	3	1	0	1	0	0	4
9106	PF8443_8_7_U447200	A	1	8	0	0	0	1	2	0	1	0	1	8
9106	PF8443_8_7_U550800	A	0	3	1	1	5	1	1	1	3	0	2	3
9106	PF8443_8_7_U525100	A	1	2	0	1	1	0	1	4	3	1	4	2
9106	PF8443_8_7_U531500	A	1	0	1	1	0	4	1	2	0	0	4	0
9106	PF8443_8_7_U522400	A	2	6	1	0	0	2	1	0	0	0	0	2
9106	PF8443_8_7_U432400	A	1	2	0	0	1	0	1	0	0	0	1	7
9106	PF8443_8_7_U536500	A	0	4	0	1	1	1	2	0	0	0	0	4
9106	PF8443_8_7_U454100	A	1	8	0	0	1	0	0	0	0	1	0	1
9106	PF8443_8_7_U520500	A	1	1	0	0	2	1	1	0	1	2	2	1
9106	PF8443_8_7_U519300	A	1	3	0	0	0	0	1	1	1	0	1	2
9106	PF8443_8_7_U446800	A	1	2	0	0	1	0	0	0	0	0	1	3
9106	PF8443_8_7_U513200	A	1	2	0	0	0	2	0	0	0	0	0	4
9106	PF8443_8_7_U513500	A	1	2	0	0	0	2	0	0	0	0	0	4
9106	PF8443_8_7_U531800	A	0	1	0	1	0	3	0	0	0	0	0	2
9106	PF8443_8_7_U561000	A	1	3	0	0	0	0	0	0	0	0	0	3
9106	PF8443_8_7_U523000	A	0	0	0	1	0	2	1	1	0	0	1	0
9106	PF8443_8_7_U435300	A	0	2	0	1	0	0	1	0	0	0	0	2
9106	PF8443_8_7_U538200	A	0	1	0	0	0	0	2	0	0	0	0	3
9106	PF8443_8_7_U486200	A	0	0	0	0	0	0	1	0	3	0	1	0
9106	PF8443_8_7_U502300	B	9	28	4	255	96	56	13	31	415	3	356	66
9106	PF8443_8_7_U499600	B	14	60	37	64	32	62	29	31	133	1	641	141
9106	PF8443_8_7_U458200	B	160	115	8	66	77	41	67	44	56	0	110	332
9106	PF8443_8_7_U432500	B	42	183	9	16	12	10	86	72	169	0	23	390
9106	PF8443_8_7_U436300	B	20	170	1	16	28	6	110	4	7	0	36	464
9106	PF8443_8_7_U445900	B	29	154	4	9	17	15	82	21	10	3	13	391
9106	PF8443_8_7_U555300	B	102	51	15	9	4	5	57	163	79	0	12	175
9106	PF8443_8_7_U446900	B	47	32	24	17	29	24	75	35	72	0	73	176
9106	PF8443_8_7_U435400	B	106	79	6	28	37	38	51	3	3	0	17	211
9106	PF8443_8_7_U460900	B	24	82	2	4	10	10	75	8	4	0	67	261
9106	PF8443_8_7_U434100	B	12	104	0	11	6	17	68	0	0	1	23	248
9106	PF8443_8_7_U518800	B	39	103	3	12	18	10	38	20	32	0	31	163
9106	PF8443_8_7_U554400	B	54	67	7	7	7	5	43	19	26	0	69	129
9106	PF8443_8_7_U437800	B	43	69	7	21	13	17	28	33	28	3	30	125
9106	PF8443_8_7_U519500	B	38	80	7	13	11	12	30	19	23	0	25	153
9106	PF8443_8_7_U534800	B	13	71	1	10	11	9	32	5	15	0	34	136
9106	PF8443_8_7_U561100	B	9	52	1	7	4	8	26	15	4	0	26	136
9106	PF8443_8_7_U555200	B	22	60	1	3	6	5	34	5	14	0	10	125
9106	PF8443_8_7_U542000	B	5	35	1	15	13	7	24	4	3	0	19	90
9106	PF8443_8_7_U561700	B	7	60	2	7	2	2	53	3	0	0	5	56
9106	PF8443_8_7_U561900	B	6	16	1	40	34	28	13	1	0	0	17	5
9106	PF8443_8_7_U449400	B	7	16	2	6	2	3	8	3	7	0	5	32
9106	PF8443_8_7_U526900	B	1	8	0	8	3	4	14	0	10	0	8	35
9106	PF8443_8_7_U444000	B	1	9	1	0	0	0	2	0	0	0	10	9
9106	PF8443_8_7_U486700	B	0	5	0	2	0	0	1	0	7	0	0	8
9106	PF8443_8_7_U439400	B	1	9	0	1	2	1	0	0	0	0	2	8
9106	PF8443_8_7_U470300	B	0	9	0	0	0	0	1	2	1	0	3	7
9106	PF8443_8_7_U550900	B	0	2	0	2	1	1	1	1	1	0	2	3
9106	PF8443_8_7_U551100	B	0	7	0	0	0	0	0	0	0	0	0	5
9106	PF8443_8_7_U503300	B	0	3	0	0	0	0	1	0	0	0	0	3
9106	PF8443_8_7_U535900	B	1	0	0	0	0	1	0	1	1	0	0	1
9106	PF8443_8_7_U443900	B?	84	145	44	227	294	143	67	124	134	0	108	292

Isolate	Sequence Name	A/B	AC_R	AC_T	AC_G	CU_R	CU_LR-ET	CU_MT	CU_S	CU_G	HB_R	HB_LR-ET	HB_T	HB_S	HB_G
10668	PF8443_8_8_U517800	A	101	0	66	3747	4385	1098	1620	43	15	51	20	110	30
10668	PF8443_8_8_U293700	A	146	1	26	33	78	2173	240	76	161	869	809	493	86
10668	PF8443_8_8_U534400	A	22	0	39	11	29	436	313	49	38	365	258	302	44
10668	PF8443_8_8_U533400	A	234	0	46	17	82	255	635	17	37	54	27	158	20
10668	PF8443_8_8_U532600	A	28	0	40	1	4	118	659	17	9	2	2	44	18
10668	PF8443_8_8_U537800	A	173	1	27	27	47	73	202	35	48	62	34	127	24
10668	PF8443_8_8_U526900	A	64	0	47	5	38	136	350	16	16	18	9	121	22
10668	PF8443_8_8_U535600	A	153	0	39	11	49	60	95	22	53	103	52	88	22
10668	PF8443_8_8_U498600	A	34	0	73	2	1	3	6	44	89	194	74	143	39
10668	PF8443_8_8_U531900	A	149	0	28	11	37	26	131	15	21	39	22	120	23
10668	PF8443_8_8_U553200	A	12	0	118	8	10	6	16	120	5	12	2	39	114
10668	PF8443_8_8_U532400	A	58	0	46	0	1	8	46	2	11	73	46	130	30
10668	PF8443_8_8_U500800	A	75	0	55	0	0	2	1	48	10	83	76	24	65
10668	PF8443_8_8_U528100	A	33	0	45	1	6	20	87	29	11	73	27	34	39
10668	PF8443_8_8_U526600	A	37	0	45	1	8	22	49	36	16	107	20	37	24
10668	PF8443_8_8_U542200	A	20	1	119	1	3	6	14	105	2	5	4	16	84
10668	PF8443_8_8_U488600	A	57	0	30	8	28	39	38	12	20	36	25	44	22
10668	PF8443_8_8_U532300	A	9	0	80	1	1	6	9	58	5	86	44	20	39
10668	PF8443_8_8_U063400	A	11	0	87	3	1	10	2	54	8	60	18	17	52
10668	PF8443_8_8_U529600	A	26	0	30	2	7	18	96	35	9	25	8	18	34
10668	PF8443_8_8_U535300	A	21	0	61	0	4	7	5	69	0	9	4	14	65
10668	PF8443_8_8_U524200	A	8	0	84	2	0	2	7	54	1	2	2	17	73
10668	PF8443_8_8_U491300	A	25	0	89	0	0	2	4	59	1	2	1	7	50
10668	PF8443_8_8_U524900	A	48	0	30	2	2	17	20	12	13	37	11	31	16
10668	PF8443_8_8_U501700	A	59	0	25	2	7	16	35	11	5	35	9	20	10
10668	PF8443_8_8_U534500	A	7	0	85	0	0	1	3	32	11	37	7	23	27
10668	PF8443_8_8_U534600	A	20	0	15	10	4	34	52	14	9	22	9	28	14
10668	PF8443_8_8_U525300	A	13	0	10	0	0	1	8	2	15	101	33	37	4
10668	PF8443_8_8_U566900	A	46	0	28	2	8	18	15	18	11	8	4	33	27
10668	PF8443_8_8_U539800	A	25	0	53	0	4	5	8	33	7	16	7	28	30
10668	PF8443_8_8_U532900	A	69	1	12	2	7	12	28	12	12	15	11	21	11
10668	PF8443_8_8_U491200	A	0	0	64	0	0	0	0	86	0	1	0	7	55
10668	PF8443_8_8_U096900	A	0	0	82	0	1	7	1	44	3	2	3	5	52
10668	PF8443_8_8_U491100	A	2	4	64	1	0	6	7	45	3	9	8	9	37
10668	PF8443_8_8_U572400	A	10	0	85	0	2	1	7	35	0	6	2	3	42
10668	PF8443_8_8_U056600	A	1	0	62	0	0	0	3	43	0	9	4	1	66
10668	PF8443_8_8_U570200	A	18	0	36	0	0	0	0	25	4	37	21	12	34
10668	PF8443_8_8_U550400	A	5	0	38	0	0	1	2	22	4	44	30	16	23
10668	PF8443_8_8_U563500	A	6	0	63	4	2	0	7	1	1	11	8	11	63
10668	PF8443_8_8_U565800	A	3	0	11	3	3	9	6	5	8	36	18	60	9
10668	PF8443_8_8_U553500	A	1	0	7	0	1	6	4	6	11	70	22	16	13
10668	PF8443_8_8_U506000	A	7	0	5	1	0	2	3	6	5	60	38	23	4
10668	PF8443_8_8_U501600	A	7	0	46	1	0	2	2	31	5	6	0	10	41
10668	PF8443_8_8_U505800	A	6	0	40	0	0	3	3	30	3	18	6	8	31
10668	PF8443_8_8_U532100	A	11	0	29	0	0	5	3	31	4	4	3	12	42
10668	PF8443_8_8_U490100	A	1	0	50	0	0	5	2	30	2	11	10	8	23
10668	PF8443_8_8_U500900	A	19	0	34	0	2	2	12	12	7	24	11	8	11
10668	PF8443_8_8_U537600	A	15	0	31	1	1	9	14	13	8	18	9	9	10
10668	PF8443_8_8_U541100	A	3	0	45	0	0	0	2	30	0	2	2	13	28
10668	PF8443_8_8_U126100	A	2	0	45	0	0	2	2	21	0	5	0	15	30
10668	PF8443_8_8_U293600	A	21	0	23	0	1	8	12	8	4	11	6	12	12
10668	PF8443_8_8_U563700	A	11	0	37	3	0	1	2	20	0	1	3	18	18
10668	PF8443_8_8_U530500	A	8	0	35	0	1	2	7	21	5	2	0	12	20
10668	PF8443_8_8_U487800	A	1	0	44	1	2	2	1	17	1	9	1	13	15
10668	PF8443_8_8_U519000	A	12	0	35	0	0	5	6	15	0	2	0	2	28
10668	PF8443_8_8_U572300	A	3	0	39	0	2	2	3	17	1	11	8	4	12
10668	PF8443_8_8_U515800	A	1	1	23	3	10	6	17	20	0	1	0	10	9
10668	PF8443_8_8_U563400	A	2	0	42	0	0	0	0	0	0	16	2	5	24
10668	PF8443_8_8_U574400	A	0	0	42	1	0	1	1	16	1	5	5	2	17
10668	PF8443_8_8_U537200	A	9	0	20	0	1	1	11	12	0	9	2	4	16
10668	PF8443_8_8_U564500	A	3	0	37	0	1	1	0	18	1	4	0	0	19
10668	PF8443_8_8_U508300	A	3	0	20	0	0	2	5	4	0	2	4	32	10
10668	PF8443_8_8_U534900	A	2	0	48	4	1	1	4	11	0	1	0	1	8
10668	PF8443_8_8_U527800	A	13	0	17	2	1	1	2	8	3	12	1	4	11
10668	PF8443_8_8_U536900	A	12	0	14	1	3	1	11	3	7	5	5	5	4
10668	PF8443_8_8_U533800	A	7	0	12	1	0	0	3	14	1	5	0	10	16
10668	PF8443_8_8_U559000	A	1	0	34	0	2	0	2	14	0	0	0	4	11

Isolate	Sequence Name	A/B	AC_R	AC_T	AC_G	CU_R	CU_LR-ET	CU_MT	CU_S	CU_G	HB_R	HB_LR-ET	HB_T	HB_S	HB_G
10668	PF8443_8_8_U520400	A	0	0	24	0	2	2	4	13	2	1	1	1	7
10668	PF8443_8_8_U066400	A	0	0	16	0	4	3	0	13	2	3	1	2	12
10668	PF8443_8_8_U510500	A	0	0	8	0	0	0	0	12	1	4	2	0	13
10668	PF8443_8_8_U532200	A	3	0	14	0	0	0	0	5	2	2	2	0	11
10668	PF8443_8_8_U548600	A	0	0	9	0	0	0	5	4	0	5	1	9	7
10668	PF8443_8_8_U067400	A	1	0	22	0	0	1	1	6	0	0	1	1	4
10668	PF8443_8_8_U491800	A	1	0	11	0	0	4	9	4	1	0	0	1	7
10668	PF8443_8_8_U540600	A	1	0	3	1	0	2	6	3	1	2	1	14	3
10668	PF8443_8_8_U510600	A	0	0	7	0	0	1	0	10	0	9	5	1	4
10668	PF8443_8_8_U097000	A	3	2	13	0	0	0	0	8	1	1	1	2	4
10668	PF8443_8_8_U530300	A	1	0	4	0	0	1	0	4	0	2	1	3	5
10668	PF8443_8_8_U510300	A	0	0	13	0	1	0	1	3	1	0	0	0	1
10668	PF8443_8_8_U529500	A	0	0	8	0	0	0	5	2	0	0	0	2	2
10668	PF8443_8_8_U536500	A	2	0	2	0	1	1	1	5	1	0	0	2	5
10668	PF8443_8_8_U540900	A	0	0	0	0	0	2	0	1	1	6	2	6	1
10668	PF8443_8_8_U097300	A	0	0	13	0	0	0	0	3	0	0	0	0	3
10668	PF8443_8_8_U527200	A	0	0	4	1	0	0	1	3	1	2	1	0	5
10668	PF8443_8_8_U536800	A	0	0	7	0	0	0	0	2	0	2	1	1	3
10668	PF8443_8_8_U541000	A	1	0	7	0	0	0	1	3	0	0	0	1	1
10668	PF8443_8_8_U572200	A	0	0	4	0	0	1	2	1	0	0	0	2	4
10668	PF8443_8_8_U532800	A	0	0	1	0	1	1	6	0	1	1	0	0	1
10668	PF8443_8_8_U553400	A	0	0	2	0	0	1	0	4	0	0	0	0	4
10668	PF8443_8_8_U533300	A	0	0	1	1	0	0	1	0	1	0	1	2	1
10668	PF8443_8_8_U560700	A	0	0	3	0	0	1	0	1	0	0	0	2	1
10668	PF8443_8_8_U550300	A	0	0	5	0	0	1	0	1	0	0	0	0	0
10668	PF8443_8_8_U497400	A	0	0	1	1	1	0	0	1	0	0	0	0	2
10668	PF8443_8_8_U521000	A	0	0	1	0	0	0	0	3	0	0	0	1	1
10668	PF8443_8_8_U501900	A	0	0	2	1	0	0	0	0	0	0	0	1	2
10668	PF8443_8_8_U521100	B	186	0	44	1	0	15	35	28	77	415	234	807	24
10668	PF8443_8_8_U505900	B	267	0	88	1	5	3	3	114	62	32	21	98	133
10668	PF8443_8_8_U489900	B	48	0	95	1	7	11	32	116	3	3	4	262	132
10668	PF8443_8_8_U574500	B	47	0	139	4	2	9	8	98	26	33	22	92	102
10668	PF8443_8_8_U522100	B	71	6	70	0	2	17	13	59	12	126	32	110	52
10668	PF8443_8_8_U500600	B	34	0	95	0	3	7	17	111	8	52	35	84	116
10668	PF8443_8_8_U564300	B	76	0	8	0	0	5	8	7	27	54	80	192	7
10668	PF8443_8_8_U512100	B	19	0	78	0	4	7	7	61	6	43	19	119	60
10668	PF8443_8_8_U559100	B	21	0	64	2	4	6	10	111	8	29	11	49	93
10668	PF8443_8_8_U528900	B	6	0	129	0	6	7	7	84	1	5	4	45	109
10668	PF8443_8_8_U488500	B	13	0	83	1	1	4	16	77	3	14	6	56	75
10668	PF8443_8_8_U529300	B	12	0	108	0	5	3	11	49	3	13	9	22	69
10668	PF8443_8_8_U574300	B	60	0	45	1	1	8	11	34	13	26	13	51	33
10668	PF8443_8_8_U533900	B	31	0	77	0	5	15	8	54	5	0	2	27	51
10668	PF8443_8_8_U511600	B	27	0	65	0	1	0	4	39	8	42	26	11	47
10668	PF8443_8_8_U553000	B	13	0	93	0	1	2	5	34	3	14	2	8	58
10668	PF8443_8_8_U523300	B	19	0	96	0	0	6	1	44	2	2	1	7	49
10668	PF8443_8_8_U506900	B	24	0	56	0	0	2	4	39	5	7	4	27	49
10668	PF8443_8_8_U563600	B	20	2	96	0	0	0	0	0	2	2	1	32	59
10668	PF8443_8_8_U540700	B	7	0	52	0	2	4	7	32	2	2	1	54	30
10668	PF8443_8_8_U528700	B	4	0	74	0	0	0	0	57	2	3	1	7	40
10668	PF8443_8_8_U493000	B	4	0	46	0	4	2	2	51	2	3	2	6	51
10668	PF8443_8_8_U507200	B	45	0	27	0	0	2	4	22	10	12	5	27	15
10668	PF8443_8_8_U566700	B	2	0	56	0	1	1	4	30	1	11	0	17	40
10668	PF8443_8_8_U563800	B	9	0	33	4	6	1	14	21	1	2	0	4	32
10668	PF8443_8_8_U490000	B	7	3	12	3	1	1	0	35	1	0	1	2	51
10668	PF8443_8_8_U559200	B	4	1	27	0	0	3	3	28	1	2	0	22	24
10668	PF8443_8_8_U563900	B	1	0	12	5	2	10	1	5	0	1	0	1	17
10668	PF8443_8_8_U550600	B	0	0	7	0	0	0	0	19	0	0	0	4	21
10668	PF8443_8_8_U549400	B	0	0	6	0	2	2	1	11	0	0	0	0	22
10668	PF8443_8_8_U521900	B	1	0	14	0	0	0	0	0	0	1	2	3	12
10668	PF8443_8_8_U484100	B	2	0	2	0	0	4	9	2	0	0	3	6	1
10668	PF8443_8_8_U501800	B	0	0	0	0	0	0	0	0	2	4	0	7	1
10668	PF8443_8_8_U497500	B	0	0	4	0	0	0	0	1	0	0	0	2	4
10668	PF8443_8_8_U097200	B	0	0	1	0	0	0	1	1	0	0	0	0	5
10668	PF8443_8_8_U520200	B	0	0	0	0	0	0	1	1	2	0	0	1	1
10668	PF8443_8_8_U484000	B?	47	2	85	10	10	45	27	80	10	54	41	65	95
10668	PF8443_8_8_U567200	B?	0	0	16	1	0	3	1	8	0	1	2	3	12
10668	PF8443_8_8_U532700	B?	0	0	3	0	0	0	0	1	0	0	0	0	4
10668	PF8443_8_8_U505500	B?	0	0	0	0	0	0	0	5	0	0	0	0	1

Isolate	Gene ID	A/B	R+R	R+LRET	R+T	R+S	R+G	R-R	R-LRET	R-T	R-S	R-G
SA075	PFSAO75_PN48	A	2438	291	141	489	82	3102	581	571	2256	67
SA075	PFSAO75_128200	A	32	869	1071	687	52	2	26	33	5	42
SA075	PFSAO75_PN57	A	48	598	684	534	22	11	96	138	18	20
SA075	PFSAO75_PN56	A	30	226	356	362	22	3	12	47	19	15
SA075	PFSAO75_PN28	A	151	1	11	35	40	319	2	6	301	43
SA075	PFSAO75_071800	A	0	2	98	62	5	9	228	302	9	6
SA075	PFSAO75_PN07	A	1	91	225	67	74	8	4	5	3	79
SA075	PFSAO75_154500	A	0	4	6	3	29	3	194	154	2	22
SA075	PFSAO75_PN52	A	68	41	47	89	124	8	8	8	10	13
SA075	PFSAO75_415400	A	9	99	161	69	11	0	1	1	0	12
SA075	PFSAO75_PN46	A	24	19	75	59	22	2	72	59	4	24
SA075	PFSAO75_PN41	A	8	87	132	48	16	7	6	6	2	11
SA075	PFSAO75_PN50	A	25	29	44	31	46	6	26	30	24	34
SA075	PFSAO75_071500	A	5	22	31	38	117	7	1	3	1	67
SA075	PFSAO75_494200	A	24	9	11	18	60	39	16	8	26	44
SA075	PFSAO75_185100	A	6	5	11	13	21	5	34	99	31	21
SA075	PFSAO75_PN24	A	11	16	11	14	91	4	1	6	2	78
SA075	PFSAO75_061400	A	0	1	3	1	66	5	11	7	0	101
SA075	PFSAO75_PN33	A	13	0	3	19	65	5	5	6	9	50
SA075	PFSAO75_PN22	A	14	10	8	4	30	7	24	21	4	40
SA075	PFSAO75_302500	A	0	8	21	13	10	0	44	46	4	12
SA075	PFSAO75_347900	A	0	8	21	13	10	0	44	46	4	12
SA075	PFSAO75_030300	A	9	8	7	9	54	12	3	7	3	45
SA075	PFSAO75_PN43	A	1	0	92	14	20	1	0	0	3	15
SA075	PFSAO75_PN53	A	31	9	15	10	31	8	4	6	1	30
SA075	PFSAO75_030700	A	5	3	7	2	49	11	3	0	3	60
SA075	PFSAO75_127900	A	1	31	37	20	28	2	3	2	1	13
SA075	PFSAO75_493800	A	7	8	9	14	23	2	8	35	8	23
SA075	PFSAO75_302300	A	10	2	3	7	36	11	0	5	32	27
SA075	PFSAO75_PN55	A	7	15	34	20	24	2	7	9	1	13
SA075	PFSAO75_051600	A	45	4	7	8	21	1	2	9	2	22
SA075	PFSAO75_PN36	A	1	0	5	9	39	5	2	3	3	51
SA075	PFSAO75_PN38	A	6	1	2	6	54	1	0	0	0	43
SA075	PFSAO75_072700	A	3	2	7	0	40	5	8	5	1	41
SA075	PFSAO75_492300	A	4	3	5	3	44	2	2	0	9	36
SA075	PFSAO75_PN51	A	6	3	13	2	30	5	0	5	1	30
SA075	SAO75_PN01	A	0	1	0	2	31	3	2	1	2	50
SA075	PFSAO75_PN10	A	3	0	5	4	32	1	2	3	3	34
SA075	PFSAO75_PN29	A	7	1	4	16	29	2	0	3	2	21
SA075	PFSAO75_494100	A	4	9	4	8	24	4	1	11	3	16
SA075	PFSAO75_PN18	A	1	4	6	6	20	0	1	4	9	26
SA075	PFSAO75_PN54	A	10	2	3	4	23	4	1	0	1	27
SA075	PFSAO75_PN30	A	0	2	6	5	23	0	2	6	0	22
SA075	PFSAO75_PN14	A	0	2	0	0	28	4	0	0	1	23
SA075	PFSAO75_PN37	A	0	1	16	6	17	0	3	0	0	15
SA075	PFSAO75_PN49	A	1	0	1	1	30	2	2	1	0	17
SA075	PFSAO75_258400	A	1	1	4	5	21	1	2	0	2	16
SA075	PFSAO75_PN13	A	1	3	5	5	19	2	1	0	1	13

Isolate	Gene ID	A/B	R+R	R+LRET	R+T	R+S	R+G	R-R	R-LRET	R-T	R-S	R-G
SA075	PFSAO75_011500	A	2	1	0	0	16	1	0	0	0	23
SA075	PFSAO75_030600	A	0	2	1	0	16	0	8	11	0	3
SA075	PFSAO75_PN04	A	0	0	1	0	24	2	0	0	0	14
SA075	PFSAO75_PN21	A	0	0	2	0	11	0	0	0	2	24
SA075	PFSAO75_493600	A	1	1	3	1	20	1	0	1	1	10
SA075	PFSAO75_PN35	A	0	0	0	0	10	2	0	0	0	27
SA075	PFSAO75_030400	A	0	1	2	0	13	0	1	0	0	14
SA075	PFSAO75_052900	A	0	1	1	1	11	1	0	0	0	14
SA075	PFSAO75_PN05	A	0	2	1	3	20	0	0	0	1	2
SA075	PFSAO75_492900	A	0	0	0	0	11	0	0	1	0	9
SA075	PFSAO75_492600	A	0	0	0	0	5	0	2	0	0	10
SA075	PFSAO75_128600	A	0	1	1	2	4	0	0	0	3	3
SA075	PFSAO75_PN26	A	2	0	0	2	4	0	0	0	1	4
SA075	PFSAO75_PN32	A	1	2	0	0	3	0	0	2	2	2
SA075	PFSAO75_PN27	A	2	0	0	0	2	0	0	0	0	7
SA075	PFSAO75_302400	A	0	0	0	0	1	0	0	0	0	5
SA075	PFSAO75_258500	A trunc	0	2	1	0	6	2	0	1	0	1
SA075	PFSAO75_PN19	B	15	6	62	43	107	28	352	528	50	102
SA075	PFSAO75_220500	B	6	88	168	67	31	1	42	81	5	39
SA075	PFSA075_PN02	B	7	0	0	13	90	7	2	7	60	75
SA075	PFSAO75_101100	B	7	3	2	5	111	12	16	19	3	78
SA075	PFSAO75_494000	B	46	0	10	10	49	11	4	11	57	43
SA075	PFSAO75_PN42	B	6	6	30	15	71	5	21	13	2	66
SA075	PFSAO75_PN34	B	7	10	6	6	77	11	7	7	11	83
SA075	PFSAO75_PN40	B	0	1	1	5	83	5	0	3	0	96
SA075	PFSAO75_490000	B	5	2	4	3	85	3	1	5	5	70
SA075	PFSAO75_101400	B	2	0	3	0	78	4	3	5	2	69
SA075	PFSAO75_PN15	B	0	2	1	1	32	2	49	30	1	32
SA075	PFSAO75_101300	B	39	0	11	12	42	1	0	4	1	35
SA075	PFSAO75_PN44	B	0	3	2	3	45	1	0	5	1	83
SA075	PFSAO75_PN23	B	9	0	3	8	51	5	3	5	21	24
SA075	PFSAO75_413400	B	2	0	2	1	63	3	3	3	2	48
SA075	PFSAO75_493900	B	5	3	9	11	55	3	0	1	1	36
SA075	PFSAO75_491400	B	2	2	3	1	52	4	1	1	10	34
SA075	PFSAO75_PN39	B	0	3	2	0	55	3	1	0	0	38
SA075	PFSAO75_071400	B	0	0	1	2	33	1	2	5	2	50
SA075	PFSAO75_PN45	B	3	0	2	2	44	0	2	2	0	21
SA075	PFSAO75_413900	B	0	0	0	0	31	0	0	2	3	39
SA075	PFSAO75_PN25	B	0	0	0	0	48	0	1	2	0	4
SA075	PFSAO75_PN31	B	1	1	3	1	23	2	1	0	0	22
SA075	PFSAO75_011200	B	1	0	0	0	7	0	11	11	1	4
SA075	PFSAO75_PN16	B	1	0	0	1	15	2	0	0	1	12
SA075	PFSAO75_127700	B	0	3	5	2	19	0	0	0	0	2
SA075	PFSAO75_011600	B	5	1	2	0	4	1	0	0	4	5
SA075	PFSAO75_PN12	B	0	0	2	2	9	0	0	0	0	6
SA075	PFSAO75_030900	B	5	1	2	0	3	1	0	0	3	4
SA075	PFSAO75_348600	B	1	0	0	1	7	0	0	0	0	8
SA075	PFSAO75_412900	B	0	0	0	0	3	1	0	0	0	2
SA075	PFSAO75_348000	B	0	0	0	0	3	1	0	0	0	1

Appendix Table 9.11. *Raw rif read counts from 454 amplicon sequencing of cDNA and gDNA from a patien isolate 8383 put under differential selection pressure. AC (acute rings and ex vivo trophozoites) and CU (culture-adapted) and fresh-from-the-arm patient samples that were matured in culture for a single IDC, (acute rings, and ex vivo late rings to schizonts). These isolates lacked whole genome sequence data. Therefore the data in the table was generated by de novo assembly of 454 reads. The numbers of reads that went into contig assembly were computed per sample and these numbers were used at the read count data. !: samples with limited sequence depth (>1000 reads).*

Isolate	Cluster ID	A/B	Gene ID	AC_R	AC_T	AC_G	CU_R	CU_LRET	CU_T	CU_S	CU_G
8383	454_RIF_0080	A	PF383_RIF_CU_00_G1_2_3	83.5	51.5	34	363	411	121	284	21
8383	454_RIF_0001	A	PF383_RIF_CU_00_G2_4	0	21.5	24	18	17	311	130	20
8383	454_RIF_0225	A	PF383_RIF_CU_00_RN_2_13	64	127.5	16	70	34	111	103	14
8383	454_RIF_0002	A	PF383_RIF_AC_00_R1_2_14	183	30.5	59	77	38	0	25	54
8383	454_RIF_0120	A	PF383_RIF_CU_00_G1_2_47	0	8	31	158	91	77	62	20
8383	454_RIF_0172	A	PF383_RIF_CU_00_RN_2_25	47	118	26	46	16	83	93	10
8383	454_RIF_0023	A	PF383_RIF_AC_24_T2_2_47	28.5	33	33	77	116	0	78	37
8383	454_RIF_0177	A	PF383_RIF_AC_00_GE_2_19	31.3	26.5	17	137	62	37	65	14
8383	454_RIF_0250	A	PF383_RIF_CU_00_RN_2_4	26.5	141	14	24	49	35	48	16
8383	454_RIF_0209	A	PF383_RIF_CU_18_TR_2_1	26.5	21.5	21	95	108	16	53	9
8383	454_RIF_0189	A	PF383_RIF_AC_24_T2_2_38	8.5	138	22	23	11	32	59	25
8383	454_RIF_0211	A	PF383_RIF_CU_00_RN_2_2	23.5	18	6	28	123	20	31	0
8383	454_RIF_0181	A	PF383_RIF_AC_24_T1_2_5	22.3	138	10	0	12	13	22	16
8383	454_RIF_0323	A	PF383_RIF_AC_00_R2_2_18	28	16	7	59	62	25	29	0
8383	454_RIF_0231	A	PF383_RIF_AC_24_T1_2_24	10	74.5	26	5	0	56	13	27
8383	454_RIF_0261	A	PF383_RIF_CU_18_TR_2_5	2	6	25	6	0	45	104	15
8383	454_RIF_0142	A	PF383_RIF_AC_00_GE_2_25	10.5	13.5	31	30	11	31	41	30
8383	454_RIF_0312	A	PF383_RIF_AC_24_T2_2_19	14	106	7	0	0	21	9	18
8383	454_RIF_0158	A	PF383_RIF_CU_00_G1_2_57	10.5	19	44	0	5	44	14	31
8383	454_RIF_0361	A	PF383_RIF_AC_00_GE_2_16	5.5	0	29	28	56	0	24	15
8383	454_RIF_0148	A	PF383_RIF_CU_00_RN_2_15	15.5	42	17	21	19	0	18	19
8383	454_RIF_0123	A	PF383_RIF_AC_24_T1_2_33	0	18	0	0	0	116	0	7
8383	454_RIF_0178	A	PF383_RIF_AC_24_T1_2_3	19.5	27.5	11	8	10	45	10	8
8383	454_RIF_0214	A	PF383_RIF_AC_00_GE_2_13	31.5	15.5	14	21	17	4	9	12
8383	454_RIF_0364	A	PF383_RIF_CU_30_SZ_2_3	7	11	7	11	0	13	58	7
8383	454_RIF_0440	A	PF383_RIF_CU_30_SZ_2_7	0.0	16.5	15	0	0	53	8	13
8383	454_RIF_0430	A	PF383_RIF_AC_24_T2_2_9	0.0	18.5	7	0	0	39	33	7
8383	454_RIF_0234	A	PF383_RIF_CU_00_G1_2_58	0	7	40	0	0	15	0	36
8383	454_RIF_0368	A	PF383_RIF_AC_24_T1_2_11	0	21	0	8	0	34	24	5
8383	454_RIF_0629	A	PF383_RIF_AC_24_T2_2_37	0	9	32	0	4	0	16	28
8383	454_RIF_0522	A	PF383_RIF_AC_00_GE_2_38	0	6	45	0	0	0	2	30
8383	454_RIF_0712	A	PF383_RIF_CU_00_G1_2_8	0	69	5	0	0	0	0	7
8383	454_RIF_0434	A	PF383_RIF_CU_00_G1_2_10	0.0	18	0	0	4	18	20	13
8383	454_RIF_0455	A	PF383_RIF_CU_30_SZ_2_17	0.0	12	16	0	0	16	10	18
8383	454_RIF_0304	A	PF383_RIF_AC_24_T1_2_2	7	18	10	10	0	9	7	10
8383	454_RIF_0265	A	PF383_RIF_CU_00_G1_2_29	0	6.5	14	15	0	14	8	13
8383	454_RIF_0422	A	PF383_RIF_CU_30_SZ_2_1	0	7	12	0	8	0	27	13
8383	454_RIF_0521	A	PF383_RIF_AC_24_T2_2_26	0	14	15	0	0	0	22	12
8383	454_RIF_0946	A	PF383_RIF_AC_00_GE_2_50	0	0	13	18	10	0	0	21
8383	454_RIF_0429	A	PF383_RIF_AC_24_T2_2_7	0.0	26	15	0	0	4	7	10
8383	454_RIF_0010	A	PF383_RIF_CU_00_G1_2_56	0	10	13	0	0	20	9	9
8383	454_RIF_0180	A	PF383_RIF_CU_18_TR_2_9	0	0	29	3	0	13	0	13
8383	454_RIF_0538	A	PF383_RIF_AC_24_T2_2_44	0	23	7	9	10	0	8	0
8383	454_RIF_1610	A	PF383_RIF_CU_30_SZ_2_57	0	0	0	0	0	0	56	0
8383	454_RIF_0787	A	PF383_RIF_CU_30_SZ_2_37	0	0	26	0	0	0	5	23
8383	454_RIF_0635	A	PF383_RIF_CU_00_G1_2_50	0	0	16	0	0	8	0	28
8383	454_RIF_0543	A	PF383_RIF_AC_00_R2_2_20	5.5	0	8	11	0	0	27	0
8383	454_RIF_0013	A	PF383_RIF_AC_00_GE_2_33	0	9	13	0	0	0	7	22
8383	454_RIF_0438	A	PF383_RIF_AC_00_R2_2_10	21.5	3	10	0	9	0	0	6

Isolate	Cluster ID	A/B	Gene ID	AC_R	AC_T	AC_G	CU_R	CU_LRET	CU_T	CU_S	CU_G
8383	454_RIF_0403	A	PF383_RIF_AC_24_T2_2_52	0	5	18	0	6	0	0	19
8383	454_RIF_0754	A	PF383_RIF_AC_00_GE_2_48	0	5	34	0	0	0	0	7
8383	454_RIF_0576	A	PF383_RIF_CU_00_G1_2_12	5	0	14	0	11	0	0	13
8383	454_RIF_0658	A	PF383_RIF_CU_00_G1_2_65	0	11	13	0	0	6	0	13
8383	454_RIF_0702	A	PF383_RIF_CU_00_G1_2_9	0	0	20	0	0	8	0	13
8383	454_RIF_0772	A	PF383_RIF_CU_00_G1_2_53	0	14	0	0	0	9	0	18
8383	454_RIF_0385	A	PF383_RIF_AC_00_GE_2_30	3	7	17	0	3	0	0	10
8383	454_RIF_1223	A	PF383_RIF_AC_24_T2_2_39	0	40	0	0	0	0	0	0
8383	454_RIF_0436	A	PF383_RIF_AC_00_GE_2_11	0.0	7	12	0	0	0	3	17
8383	454_RIF_0424	A	PF383_RIF_AC_24_T2_2_3	11.5	10.5	8	0	0	0	0	8
8383	454_RIF_0496	A	PF383_RIF_CU_00_G1_2_4	0.0	16	8	3	0	0	0	11
8383	454_RIF_0510	A	PF383_RIF_CU_00_G1_2_14	7	6	10	0	5	0	0	9
8383	454_RIF_0005	A	PF383_RIF_AC_00_GE_2_32	0	0	24	0	0	0	0	12
8383	454_RIF_0513	A	PF383_RIF_AC_24_T2_2_13	11	9	12	0	0	0	0	0
8383	454_RIF_0333	A	PF383_RIF_CU_00_G1_2_60	0	7	8	0	0	2	0	13
8383	454_RIF_0499	A	PF383_RIF_CU_00_G1_2_5	0.0	6	9	0	7	0	0	8
8383	454_RIF_0911	A	PF383_RIF_AC_24_T2_2_35	0	13	0	0	0	0	0	14
8383	454_RIF_0520	A	PF383_RIF_AC_24_T2_2_24	1	9.5	7	0	0	0	0	9
8383	454_RIF_0770	A	PF383_RIF_AC_24_T1_2_28	0	21	0	0	0	0	0	5
8383	454_RIF_0505	A	PF383_RIF_AC_00_GE_2_2	0.0	5	9	5	0	0	4	0
8383	454_RIF_0693	A	PF383_RIF_AC_00_GE_2_10	0	0	11	0	0	0	0	10
8383	454_RIF_1050	A	PF383_RIF_AC_00_GE_2_91	0	0	10	0	0	0	0	9
8383	454_RIF_0939	A	PF383_RIF_AC_00_GE_2_49	0	0	16	0	0	0	0	2
8383	454_RIF_0771	A	PF383_RIF_AC_00_R1_2_18	11.5	6	0	0	0	0	0	0
8383	454_RIF_0857	A	PF383_RIF_AC_00_GE_2_21	0	0	10	0	0	0	0	7
8383	454_RIF_0854	A	PF383_RIF_CU_00_G1_2_13	0	0	9	0	0	0	0	7
8383	454_RIF_1446	A	PF383_RIF_AC_00_R1_2_31	11	0	0	0	0	0	0	0
8383	454_RIF_1178	A	PF383_RIF_AC_00_R1_2_12	10	0	0	0	0	0	0	0
8383	454_RIF_1579	A	PF383_RIF_CU_00_G2_19	0	0	0	0	0	0	0	10
8383	454_RIF_0495	A	PF383_RIF_AC_00_GE_2_3	0.0	0	9	0	0	0	0	0
8383	454_RIF_0900	A	PF383_RIF_CU_00_G1_2_32	0	0	0	0	0	0	0	9
8383	454_RIF_1062	A	PF383_RIF_CU_00_G2_21	0	0	0	0	0	0	0	8
8383	454_RIF_1230	A	PF383_RIF_AC_00_GE_2_44	0	0	8	0	0	0	0	0
8383	454_RIF_1487	A	PF383_RIF_AC_00_GE_2_83	0	0	8	0	0	0	0	0
8383	454_RIF_1224	A	PF383_RIF_CU_00_G1_2_44	0	0	0	0	0	0	0	7
8383	454_RIF_1497	A	PF383_RIF_CU_06_PR_2_29	0	0	0	0	7	0	0	0
8383	454_RIF_0126	A	PF383_RIF_AC_00_GE_2_54	0	0	6	0	0	0	0	0
8383	454_RIF_1245	A	PF383_RIF_CU_00_G1_2_46	0	0	0	0	0	0	0	6
8383	454_RIF_1467	A	PF383_RIF_AC_00_R2_2_29	6	0	0	0	0	0	0	0
8383	454_RIF_0531	A	PF383_RIF_CU_00_G2_27	0	0	0	0	0	0	0	5
8383	454_RIF_1671	A	PF383_RIF_AC_00_R2_2_39	5	0	0	0	0	0	0	0
8383	454_RIF_1091	A	PF383_RIF_CU_00_G2_1	0	0	0	0	0	0	0	4
8383	454_RIF_1465	A	PF383_RIF_AC_00_GE_2_79	0	0	4	0	0	0	0	0
8383	454_RIF_0476	A	PF383_RIF_CU_30_SZ_2_68	0.0	0	0	0	0	0	2	0
8383	454_RIF_1686	A	PF383_RIF_CU_18_TR_2_50	0	0	0	0	0	2	0	0
8383	454_RIF_1894	A	PF383_RIF_CU_00_RN_2_53	0	0	0	2	0	0	0	0
8383	454_RIF_1870	AX	PF383_RIF_AC_24_T1_2_59	0	13	0	0	0	0	0	0
8383	454_RIF_1031	AX	PF383_RIF_AC_00_GE_2_81	0	0	8	0	0	0	3	0
8383	454_RIF_1411	AX	PF383_RIF_CU_00_RN_2_34	0	0	0	11	0	0	0	0

Isolate	Cluster ID	A/B	Gene ID	AC_R	AC_T	AC_G	CU_R	CU_LRET	CU_T	CU_S	CU_G
8383	454_RIF_1520	AX	PF383_RIF_CU_18_TR_2_43	0	0	0	0	0	11	0	0
8383	454_RIF_1919	AX	PF383_RIF_AC_00_GE_2_99	0	0	8	0	0	0	0	0
8383	454_RIF_1684	AX	PF383_RIF_AC_24_T1_2_51	0	6	0	0	0	0	0	0
8383	454_RIF_0171	B	PF383_RIF_AC_24_T1_2_42	29.5	50	42	119	94	31	259	58
8383	454_RIF_0012	B	PF383_RIF_AC_24_T1_2_15	0	14.5	59	20	40	23	103	44
8383	454_RIF_0275	B	PF383_RIF_AC_24_T2_2_31	18.5	35	0	51	49	34	69	23
8383	454_RIF_0286	B	PF383_RIF_AC_24_T1_2_39	0	12	34	14	12	9	164	13
8383	454_RIF_0163	B	PF383_RIF_AC_24_T1_2_36	6	22.5	38	21	35	6	16	52
8383	454_RIF_0143	B	PF383_RIF_CU_30_SZ_2_13	16.5	42	23	0	8	40	47	13
8383	454_RIF_0196	B	PF383_RIF_AC_24_T2_2_57	13.5	13.5	49	24	0	0	19	29
8383	454_RIF_0282	B	PF383_RIF_AC_00_R1_2_17	16.5	13	22	47	15	0	12	20
8383	454_RIF_0222	B	PF383_RIF_AC_00_GE_2_27	10	12	17	41	5	10	12	23
8383	454_RIF_0128	B	PF383_RIF_CU_00_G1_2_69	5	8	37	14	5	10	14	36
8383	454_RIF_0324	B	PF383_RIF_AC_24_T1_2_22	0	39.5	18	7	17	0	38	7
8383	454_RIF_0675	B	PF383_RIF_AC_00_GE_2_68	0	36.5	46	0	0	0	0	43
8383	454_RIF_0116	B	PF383_RIF_AC_00_GE_2_31	0	7	11	62	0	9	10	19
8383	454_RIF_0417	B	PF383_RIF_AC_24_T2_2_58	6	0	29	0	0	7	26	35
8383	454_RIF_0388	B	PF383_RIF_CU_30_SZ_2_16	6	0	21	27	20	6	11	10
8383	454_RIF_0150	B	PF383_RIF_AC_00_GE_2_45	0	0	30	13	16	12	14	14
8383	454_RIF_0321	B	PF383_RIF_AC_24_T1_2_19	0	9.5	10	10	0	27	15	20
8383	454_RIF_0491	B	PF383_RIF_AC_24_T1_2_40	16	18	21	0	0	0	0	20
8383	454_RIF_0818	B	PF383_RIF_CU_00_RN_2_38	0	0	14	10	0	0	0	13
8383	454_RIF_0161	B	PF383_RIF_CU_00_G1_2_75	0	0	28	0	0	0	0	8
8383	454_RIF_1551	B	PF383_RIF_CU_30_SZ_2_55	0	0	0	0	0	0	30	0
8383	454_RIF_0685	B	PF383_RIF_AC_24_T2_2_1	0	16	9	0	0	4	0	0
8383	454_RIF_0822	B	PF383_RIF_CU_30_SZ_2_53	0	0	7	0	0	0	9	10
8383	454_RIF_0748	B	PF383_RIF_AC_24_T2_2_33	0	7	10	0	0	5	0	0
8383	454_RIF_0736	B	PF383_RIF_AC_00_GE_2_40	0	6	7	0	0	0	0	0
8383	454_RIF_0412	B	PF383_RIF_CU_00_G1_2_77	0	0	0	0	0	0	0	8
8383	454_RIF_1044	B	PF383_RIF_AC_24_T2_2_69	0	6.5	0	0	0	0	0	0
8383	454_RIF_1400	B	PF383_RIF_CU_00_G2_9	0	0	0	0	0	0	0	6
8383	454_RIF_1501	B	PF383_RIF_CU_18_TR_2_39	0	0	0	0	0	6	0	0
8383	454_RIF_1531	B	PF383_RIF_AC_24_T1_2_45	0	5	0	0	0	0	0	0
8383	454_RIF_1809	B	PF383_RIF_CU_06_PR_2_43	0	0	0	0	5	0	0	0
8383	454_RIF_0007	B	PF383_RIF_CU_00_G1_2_80	0	0	0	0	0	0	0	3
8383	454_RIF_1801	BX	PF383_RIF_AC_00_R1_2_40	15	0	0	0	0	0	0	0
8383	454_RIF_1692	BX	PF383_RIF_CU_06_PR_2_39	0	0	0	0	11	0	0	0
8383	454_RIF_1571	BX	PF383_RIF_CU_00_G2_18	0	0	0	0	0	0	0	4
8383	454_RIF_1815	BX	PF383_RIF_AC_24_T1_2_54	0	4	0	0	0	0	0	0
8383	454_RIF_1587	XA	PF383_RIF_CU_00_RN_2_43	0	0	0	4	0	0	0	0

Isolate	Cluster	A/B	Gene ID	R	T	S	G
10727	454_RIF_0660	A	PF727_RIF_AC_22_T2_2_29	0	316.7	0	16
10727	454_RIF_0475	A	PF727_RIF_AC_22_T1_2_21	8	292	15	9
10727	454_RIF_0481	A	PF727_RIF_AC_00_G1_2_48	0	292	0	5
10727	454_RIF_0009	A	PF727_RIF_AC_00_ER_2_1	235	0.0	0	0
10727	454_RIF_0444	A	PF727_RIF_AC_22_T3_6	0	83.7	5	64
10727	454_RIF_0509	A	PF727_RIF_AC_22_T1_2_6	0	96.3	4	18
10727	454_RIF_0776	A	PF727_RIF_AC_00_G2_51	0	0	0	112
10727	454_RIF_0630	A	PF727_RIF_AC_22_T3_17	0	36.7	0	54
10727	454_RIF_0723	A	PF727_RIF_AC_22_T1_2_9	0	87.7	0	0
10727	454_RIF_0298	A	PF727_RIF_AC_00_G2_4	0	0.0	0	83
10727	454_RIF_0597	A	PF727_RIF_AC_00_G2_28	0	0.0	0	83
10727	454_RIF_0550	A	PF727_RIF_AC_00_G2_40	0	0.0	0	68
10727	454_RIF_0357	A	PF727_RIF_AC_00_G1_2_4	0	30	7	25
10727	454_RIF_0027	A	PF727_RIF_AC_00_G2_24	0	0.0	0	59
10727	454_RIF_0636	A	PF727_RIF_AC_00_G2_21	0	0.0	0	59
10727	454_RIF_0534	A	PF727_RIF_AC_22_T3_11	4	41.7	0	10
10727	454_RIF_0623	A	PF727_RIF_AC_22_T1_2_13	0	43	0	12
10727	454_RIF_0613	A	PF727_RIF_AC_22_T3_9	0	38	0	8
10727	454_RIF_0738	A	PF727_RIF_AC_22_T1_2_11	0	5.5	0	39
10727	454_RIF_0230	A	PF727_RIF_AC_00_G1_2_80	4	2	0	38
10727	454_RIF_0234	A	PF727_RIF_AC_00_G1_2_46	0	7	0	37
10727	454_RIF_0581	A	PF727_RIF_AC_22_T1_2_2	0	43	0	0
10727	454_RIF_0523	A	PF727_RIF_AC_00_G2_13	0	0.0	0	42
10727	454_RIF_0591	A	PF727_RIF_AC_00_G1_2_10	0	22	0	18
10727	454_RIF_0608	A	PF727_RIF_AC_00_G2_15	0	0.0	0	40
10727	454_RIF_0586	A	PF727_RIF_AC_00_G1_2_7	0	26.3	0	13
10727	454_RIF_0612	A	PF727_RIF_AC_00_G2_31	0	0.0	0	37
10727	454_RIF_0840	A	PF727_RIF_AC_22_T2_2_1	0	29	0	8
10727	454_RIF_1436	A	PF727_RIF_AC_00_G1_2_78	0	0	0	37
10727	454_RIF_0546	A	PF727_RIF_AC_22_T3_22	0	23.3	0	13
10727	454_RIF_0688	A	PF727_RIF_AC_00_G1_2_9	0	4.5	0	30
10727	454_RIF_0588	A	PF727_RIF_AC_22_T2_2_4	0	27.3	0	7
10727	454_RIF_0494	A	PF727_RIF_AC_00_G2_10	0	0.0	0	33
10727	454_RIF_0123	A	PF727_RIF_AC_22_T2_2_23	0	17	0	15
10727	454_RIF_0503	A	PF727_RIF_AC_00_G2_8	0	0.0	0	32
10727	454_RIF_0280	A	PF727_RIF_AC_00_G2_42	0	0.0	0	30
10727	454_RIF_0450	A	PF727_RIF_AC_22_T2_2_10	0	18	0	12
10727	454_RIF_0819	A	PF727_RIF_AC_00_G2_49	0	0	0	30
10727	454_RIF_0704	A	PF727_RIF_AC_00_G2_29	0	0	0	29
10727	454_RIF_1543	A	PF727_RIF_AC_00_G2_52	0	0	0	29
10727	454_RIF_1902	A	PF727_RIF_AC_22_T2_2_58	0	29	0	0
10727	454_RIF_0603	A	PF727_RIF_AC_00_G1_2_19	0	0.0	0	28
10727	454_RIF_0191	A	PF727_RIF_AC_22_T2_2_21	0	9.5	0	18
10727	454_RIF_0627	A	PF727_RIF_AC_00_G1_2_30	0	18.3	0	9
10727	454_RIF_0706	A	PF727_RIF_AC_00_G2_23	0	0	0	27
10727	454_RIF_0002	A	PF727_RIF_AC_00_G1_2_28	0	21.7	0	5
10727	454_RIF_0476	A	PF727_RIF_AC_00_G2_27	0	0	0	24
10727	454_RIF_0944	A	PF727_RIF_AC_00_G2_46	0	0	0	24
10727	454_RIF_1133	A	PF727_RIF_AC_00_G1_2_13	0	0	0	23
10727	454_RIF_0654	A	PF727_RIF_AC_00_G1_2_47	5	5.5	0	12
10727	454_RIF_1195	A	PF727_RIF_AC_00_G1_2_27	0	0	0	21
10727	454_RIF_0590	A	PF727_RIF_AC_22_T1_2_7	0	14.7	0	6
10727	454_RIF_0610	A	PF727_RIF_AC_00_G2_72	0	0.0	0	20
10727	454_RIF_0780	A	PF727_RIF_AC_22_T2_2_26	0	12.5	0	6
10727	454_RIF_0554	A	PF727_RIF_AC_00_G2_44	0	0.0	0	18
10727	454_RIF_0816	A	PF727_RIF_AC_00_G2_43	0	0	0	18
10727	454_RIF_0901	A	PF727_RIF_AC_00_G2_14	0	0	0	18
10727	454_RIF_0789	A	PF727_RIF_AC_00_G2_45	0	0	0	17
10727	454_RIF_0536	A	PF727_RIF_AC_00_G2_19	0	0.0	0	16
10727	454_RIF_0646	A	PF727_RIF_AC_22_T3_20	0	5	0	11
10727	454_RIF_0716	A	PF727_RIF_AC_00_G2_25	0	0	0	16
10727	454_RIF_0926	A	PF727_RIF_AC_00_G1_2_41	0	9	0	7
10727	454_RIF_1086	A	PF727_RIF_AC_00_G1_2_5	0	0	0	16

Isolate	Cluster	A/B	Gene ID	R	T	S	G
10727	454_RIF_1182	A	PF727_RIF_AC_22_T3_7	0	16	0	0
10727	454_RIF_0815	A	PF727_RIF_AC_22_T1_2_30	0	15.5	0	0
10727	454_RIF_0533	A	PF727_RIF_AC_00_G2_39	0	0.0	0	15
10727	454_RIF_0713	A	PF727_RIF_AC_00_G2_11	0	0	0	15
10727	454_RIF_0694	A	PF727_RIF_AC_00_G2_55	0	0	0	14
10727	454_RIF_0760	A	PF727_RIF_AC_00_G1_2_39	0	4	0	10
10727	454_RIF_0925	A	PF727_RIF_AC_00_G1_2_31	0	5	0	9
10727	454_RIF_0971	A	PF727_RIF_AC_00_G1_2_40	0	3	0	11
10727	454_RIF_1110	A	PF727_RIF_AC_00_G1_2_11	0	0	0	14
10727	454_RIF_1213	A	PF727_RIF_AC_00_G1_2_35	0	0	0	14
10727	454_RIF_0511	A	PF727_RIF_AC_00_G1_2_77	0	0.0	0	13
10727	454_RIF_0693	A	PF727_RIF_AC_00_G1_2_16	0	0	0	13
10727	454_RIF_0735	A	PF727_RIF_AC_00_G2_17	0	0.0	0	13
10727	454_RIF_1266	A	PF727_RIF_AC_00_G1_2_42	0	0	0	13
10727	454_RIF_0038	A	PF727_RIF_AC_00_G1_2_18	0	2	0	10
10727	454_RIF_0684	A	PF727_RIF_AC_00_G1_2_2	0	6	0	6
10727	454_RIF_0695	A	PF727_RIF_AC_00_G2_76	0	0	0	12
10727	454_RIF_0703	A	PF727_RIF_AC_00_G1_2_6	0	5	0	7
10727	454_RIF_0873	A	PF727_RIF_AC_00_G2_79	0	0	0	12
10727	454_RIF_1100	A	PF727_RIF_AC_00_G2_3	0	0	0	12
10727	454_RIF_1334	A	PF727_RIF_AC_00_G1_2_50	0	0	0	12
10727	454_RIF_0616	A	PF727_RIF_AC_22_T2_2_19	0	6	0	5
10727	454_RIF_0690	A	PF727_RIF_AC_00_G2_2	0	0	0	11
10727	454_RIF_0886	A	PF727_RIF_AC_00_G2_5	0	0	0	11
10727	454_RIF_0887	A	PF727_RIF_AC_00_G2_9	0	0	0	11
10727	454_RIF_1161	A	PF727_RIF_AC_00_G1_2_22	0	0	0	11
10727	454_RIF_1328	A	PF727_RIF_AC_00_G2_32	0	0	0	11
10727	454_RIF_0001	A	PF727_RIF_AC_00_G2_18	0	0	0	10.5
10727	454_RIF_0005	A	PF727_RIF_AC_00_G1_2_85	0	0.0	0	10
10727	454_RIF_0833	A	PF727_RIF_AC_00_G2_6	0	0	0	10
10727	454_RIF_1623	A	PF727_RIF_AC_00_G2_60	0	0	0	10
10727	454_RIF_0848	A	PF727_RIF_AC_00_G1_2_8	0	2	0	7
10727	454_RIF_1240	A	PF727_RIF_AC_00_G1_2_32	0	0	0	9
10727	454_RIF_1776	A	PF727_RIF_AC_22_T1_2_45	0	9	0	0
10727	454_RIF_0004	A	PF727_RIF_AC_00_ER_2_2	8	0.0	0	0
10727	454_RIF_0266	A	PF727_RIF_AC_00_G2_71	0	0	0	8
10727	454_RIF_0953	A	PF727_RIF_AC_00_G2_37	0	0	0	8
10727	454_RIF_1111	A	PF727_RIF_AC_00_G1_2_15	0	0	0	8
10727	454_RIF_1141	A	PF727_RIF_AC_00_G1_2_14	0	0	0	8
10727	454_RIF_0251	A	PF727_RIF_AC_00_G2_66	0	0	0	7
10727	454_RIF_0461	A	PF727_RIF_AC_22_T3_14	0	7	0	0
10727	454_RIF_1121	A	PF727_RIF_AC_00_G1_2_1	0	0	0	7
10727	454_RIF_1335	A	PF727_RIF_AC_00_G1_2_63	0	0	0	7
10727	454_RIF_1401	A	PF727_RIF_AC_22_T1_2_29	0	7	0	0
10727	454_RIF_0013	A	PF727_RIF_AC_00_G1_2_90	0	0.0	0	6
10727	454_RIF_0403	A	PF727_RIF_AC_00_G2_47	0	0	0	6
10727	454_RIF_0497	A	PF727_RIF_AC_22_T1_2_3	0	0.0	0	6
10727	454_RIF_0551	A	PF727_RIF_AC_00_G2_53	0	0.0	0	6
10727	454_RIF_0719	A	PF727_RIF_AC_00_G2_26	0	0	0	6
10727	454_RIF_0831	A	PF727_RIF_AC_00_G2_1	0	0	0	6
10727	454_RIF_1001	A	PF727_RIF_AC_00_G2_38	0	0	0	6
10727	454_RIF_1150	A	PF727_RIF_AC_00_G2_7	0	0	0	6
10727	454_RIF_1187	A	PF727_RIF_AC_22_T3_8	0	6	0	0
10727	454_RIF_1251	A	PF727_RIF_AC_22_T3_15	0	6	0	0
10727	454_RIF_1601	A	PF727_RIF_AC_00_G1_2_94	0	0	0	6
10727	454_RIF_0872	A	PF727_RIF_AC_00_G2_48	0	0	0	5
10727	454_RIF_0892	A	PF727_RIF_AC_00_G2_57	0	0	0	5
10727	454_RIF_0909	A	PF727_RIF_AC_00_G2_20	0	0	0	5
10727	454_RIF_1282	A	PF727_RIF_AC_00_G1_2_45	0	0	0	5
10727	454_RIF_0015	A	PF727_RIF_AC_00_G1_2_82	0	0.0	0	4
10727	454_RIF_0827	A	PF727_RIF_AC_00_G1_2_98	0	0	0	3
10727	454_RIF_0835	A	PF727_RIF_AC_00_G1_2_100	0	0	0	3
10727	454_RIF_1506	A	PF727_RIF_AC_00_G1_2_83	0	0	0	3

Isolate	Cluster	A/B	Gene ID	R	T	S	G
10727	454_RIF_1652	A	PF727_RIF_AC_00_G2_70	0	0	0	3
10727	454_RIF_0820	AX	PF727_RIF_AC_00_G2_50	0	0	0	27
10727	454_RIF_1602	AX	PF727_RIF_AC_00_G2_62	0	0	0	13
10727	454_RIF_1787	AX	PF727_RIF_AC_22_T3_44	0	8	0	0
10727	454_RIF_1015	AX	PF727_RIF_AC_00_G2_41	0	0	0	7
10727	454_RIF_1029	AX	PF727_RIF_AC_00_G2_63	0	0	0	6
10727	454_RIF_1037	AX	PF727_RIF_AC_00_G2_56	0	0	0	6
10727	454_RIF_1858	AX	PF727_RIF_AC_00_G2_74	0	0	0	5
10727	454_RIF_1807	AX	PF727_RIF_AC_22_T3_45	0	3	0	0
10727	454_RIF_1060	AX	PF727_RIF_AC_00_G2_73	0	0	0	2
10727	454_RIF_0043	B	PF727_RIF_AC_00_G2_16	0	0	0	134
10727	454_RIF_0487	B	PF727_RIF_AC_00_G1_2_69	0	46.7	0	70
10727	454_RIF_0890	B	PF727_RIF_AC_00_G2_12	0	0.0	0	73
10727	454_RIF_0564	B	PF727_RIF_AC_00_G1_2_60	7	18.7	0	43
10727	454_RIF_0029	B	PF727_RIF_AC_00_G1_2_74	0	9.5	0	51
10727	454_RIF_0556	B	PF727_RIF_AC_22_T1_2_22	0	21.3	0	39
10727	454_RIF_0990	B	PF727_RIF_AC_00_G1_2_87	0	0	0	48
10727	454_RIF_1075	B	PF727_RIF_AC_00_G1_2_95	0	10	0	37
10727	454_RIF_0380	B	PF727_RIF_AC_00_G1_2_23	0	19	4	23
10727	454_RIF_0668	B	PF727_RIF_AC_22_T3_25	0	8.7	0	33
10727	454_RIF_0979	B	PF727_RIF_AC_22_T1_2_24	0	5	0	36
10727	454_RIF_0810	B	PF727_RIF_AC_00_G1_2_68	4	12	0	24
10727	454_RIF_0518	B	PF727_RIF_AC_22_T2_2_43	6	5.5	0	28
10727	454_RIF_0060	B	PF727_RIF_AC_00_G2_33	0	5	0	31.5
10727	454_RIF_0237	B	PF727_RIF_AC_00_G2_30	10	9	0	14
10727	454_RIF_0621	B	PF727_RIF_AC_22_T2_2_15	12	6.5	0	14
10727	454_RIF_0786	B	PF727_RIF_AC_00_G1_2_54	0	11	0	19
10727	454_RIF_0667	B	PF727_RIF_AC_22_T2_2_33	1	9	0	19
10727	454_RIF_0673	B	PF727_RIF_AC_00_G2_34	0	0.0	0	28
10727	454_RIF_1394	B	PF727_RIF_AC_00_G1_2_67	0	0	0	25
10727	454_RIF_0614	B	PF727_RIF_AC_00_G1_2_26	0	5	0	19
10727	454_RIF_1528	B	PF727_RIF_AC_00_G1_2_86	0	0	0	24
10727	454_RIF_0753	B	PF727_RIF_AC_00_G2_64	0	0.0	0	22
10727	454_RIF_0992	B	PF727_RIF_AC_00_G1_2_62	0	6	0	16
10727	454_RIF_0321	B	PF727_RIF_AC_22_T3_23	0	6.5	0	13
10727	454_RIF_0271	B	PF727_RIF_AC_22_T1_2_16	0	12.3	0	7
10727	454_RIF_0343	B	PF727_RIF_AC_00_G1_2_61	0	0	0	19
10727	454_RIF_1347	B	PF727_RIF_AC_00_G1_2_59	0	0	0	19
10727	454_RIF_0807	B	PF727_RIF_AC_00_G2_35	0	0.0	0	18
10727	454_RIF_0648	B	PF727_RIF_AC_22_T1_2_19	0	8	0	9
10727	454_RIF_1272	B	PF727_RIF_AC_22_T3_21	0	17	0	0
10727	454_RIF_0978	B	PF727_RIF_AC_00_G1_2_55	0	6	0	9
10727	454_RIF_1402	B	PF727_RIF_AC_00_G1_2_66	0	0	0	15
10727	454_RIF_0345	B	PF727_RIF_AC_00_G1_2_88	0	0	0	14
10727	454_RIF_0007	B	PF727_RIF_AC_00_G1_2_84	0	0.0	0	12
10727	454_RIF_0728	B	PF727_RIF_AC_00_G1_2_24	0	0.0	0	12
10727	454_RIF_0986	B	PF727_RIF_AC_00_G2_36	0	0	0	12
10727	454_RIF_1388	B	PF727_RIF_AC_00_G1_2_65	0	0	0	12
10727	454_RIF_1927	B	PF727_RIF_AC_22_T1_2_52	0	11	0	0
10727	454_RIF_0069	B	PF727_RIF_AC_00_G2_22	0	0	0	10
10727	454_RIF_0165	B	PF727_RIF_AC_00_G1_2_72	0	4	0	6
10727	454_RIF_0449	B	PF727_RIF_AC_00_G2_78	0	0.0	0	9
10727	454_RIF_0805	B	PF727_RIF_AC_00_G2_58	0	0.0	0	7
10727	454_RIF_1039	B	PF727_RIF_AC_00_G2_67	0	0	0	7
10727	454_RIF_0161	B	PF727_RIF_AC_00_G1_2_73	0	0	0	5
10727	454_RIF_0342	B	PF727_RIF_AC_00_G2_68	0	0	0	5
10727	454_RIF_0395	B	PF727_RIF_AC_32_S2_10	0	0	5	0
10727	454_RIF_0726	B	PF727_RIF_AC_00_G2_61	0	0.0	0	5
10727	454_RIF_0934	B	PF727_RIF_AC_22_T3_18	0	5	0	0
10727	454_RIF_1425	B	PF727_RIF_AC_22_T1_2_28	0	5	0	0
10727	454_RIF_1431	B	PF727_RIF_AC_22_T2_2_37	0	4	0	0
10727	454_RIF_1711	B	PF727_RIF_AC_00_G2_69	0	0	0	4
10727	454_RIF_0918	B	PF727_RIF_AC_00_G2_59	0	0.0	0	3
10727	454_RIF_0143	B	PF727_RIF_AC_12_R2_15	2	0	0	0
10727	454_RIF_0412	B	PF727_RIF_AC_12_R2_10	2	0	0	0
10727	454_RIF_1019	B	PF727_RIF_AC_00_G2_75	0	0	0	2
10727	454_RIF_1713	BX	PF727_RIF_AC_22_T3_43	0	3	0	0

Isolate	Cluster	A/B	Gene ID	R	LRET	ET	G
10735	454_RIF_0001	A	PF735_RIF_AC_00_GE_2_33	0	0	0	46
10735	454_RIF_0002	A	PF735_RIF_AC_00_GE_2_26	5	0	0	14
10735	454_RIF_0013	A	PF735_RIF_AC_00_GE_2_88	0	0	0	7
10735	454_RIF_0015	A	PF735_RIF_AC_00_GE_2_61	0	0	0	13
10735	454_RIF_0023	A	PF735_RIF_AC_00_GE_2_51	0	0	0	28
10735	454_RIF_0212	A	PF735_RIF_AC_00_GE_2_9	0	0	0	11
10735	454_RIF_0266	A	PF735_RIF_AC_00_GE_2_22	0	0	0	19
10735	454_RIF_0431	A	PF735_RIF_AC_00_MR_37	3	0	0	0
10735	454_RIF_0526	A	PF735_RIF_AC_00_GE_2_32	0	0	0	19
10735	454_RIF_0535	A	PF735_RIF_AC_30_ET_10	12	57	117	15
10735	454_RIF_0547	A	PF735_RIF_AC_30_ET_12	6	8	10	18
10735	454_RIF_0578	A	PF735_RIF_AC_20_PR_1	8	19	58	27
10735	454_RIF_0620	A	PF735_RIF_AC_00_MR_6	6	27	40	13
10735	454_RIF_0625	A	PF735_RIF_AC_20_PR_4	3	4	6	32
10735	454_RIF_0643	A	PF735_RIF_AC_00_MR_9	14	10	19	15
10735	454_RIF_0650	A	PF735_RIF_AC_00_GE_2_54	0	0	0	14
10735	454_RIF_0687	A	PF735_RIF_AC_00_GE_2_2	10	0	5	11
10735	454_RIF_0695	A	PF735_RIF_AC_00_GE_2_7	0	0	0	8
10735	454_RIF_0700	A	PF735_RIF_AC_30_ET_2	0	0	9	43
10735	454_RIF_0729	A	PF735_RIF_AC_00_GE_2_24	0	0	0	20
10735	454_RIF_0739	A	PF735_RIF_AC_00_GE_2_27	7	0	2	6
10735	454_RIF_0758	A	PF735_RIF_AC_30_ET_11	0	10	20	8
10735	454_RIF_0765	A	PF735_RIF_AC_00_GE_2_37	5	0	0	32
10735	454_RIF_0838	A	PF735_RIF_AC_00_GE_2_4	7	0	0	20
10735	454_RIF_0841	A	PF735_RIF_AC_30_ET_3	0	0	6	50
10735	454_RIF_0861	A	PF735_RIF_AC_00_GE_2_14	0	0	9	19
10735	454_RIF_0881	A	PF735_RIF_AC_00_GE_2_18	0	0	14	11
10735	454_RIF_0882	A	PF735_RIF_AC_30_ET_4	0	0	7	16
10735	454_RIF_0888	A	PF735_RIF_AC_00_GE_2_17	0	0	0	23
10735	454_RIF_0889	A	PF735_RIF_AC_00_GE_2_104	0	0	0	8
10735	454_RIF_0896	A	PF735_RIF_AC_20_PR_2	0	13	22	0
10735	454_RIF_0922	A	PF735_RIF_AC_00_MR_8	7	0	0	21
10735	454_RIF_0937	A	PF735_RIF_AC_30_ET_13	0	0	17	7
10735	454_RIF_0945	A	PF735_RIF_AC_00_GE_2_41	0	0	4	35
10735	454_RIF_0956	A	PF735_RIF_AC_00_GE_2_43	11	0	0	14
10735	454_RIF_0961	A	PF735_RIF_AC_00_GE_2_47	7	0	6	30
10735	454_RIF_0982	A	PF735_RIF_AC_30_ET_16	0	0	9	23
10735	454_RIF_1011	A	PF735_RIF_AC_00_GE_2_92	0	0	10	44
10735	454_RIF_1084	A	PF735_RIF_AC_00_GE_2_1	0	7	0	10
10735	454_RIF_1088	A	PF735_RIF_AC_00_GE_2_6	0	0	0	6
10735	454_RIF_1089	A	PF735_RIF_AC_00_MR_1	20	0	0	0
10735	454_RIF_1128	A	PF735_RIF_AC_00_GE_2_10	0	0	0	10
10735	454_RIF_1131	A	PF735_RIF_AC_00_GE_2_16	0	0	0	11
10735	454_RIF_1142	A	PF735_RIF_AC_00_GE_2_5	0	0	0	13
10735	454_RIF_1149	A	PF735_RIF_AC_00_GE_2_11	0	0	0	20
10735	454_RIF_1151	A	PF735_RIF_AC_00_GE_2_13	0	0	0	20
10735	454_RIF_1156	A	PF735_RIF_AC_30_ET_5	0	0	21	0
10735	454_RIF_1158	A	PF735_RIF_AC_00_GE_2_15	0	0	0	12
10735	454_RIF_1171	A	PF735_RIF_AC_00_GE_2_19	0	0	0	20
10735	454_RIF_1180	A	PF735_RIF_AC_00_GE_2_35	0	0	0	13
10735	454_RIF_1184	A	PF735_RIF_AC_20_PR_3	0	0	0	0
10735	454_RIF_1190	A	PF735_RIF_AC_00_GE_2_25	0	0	0	40
10735	454_RIF_1197	A	PF735_RIF_AC_00_GE_2_30	0	0	0	6
10735	454_RIF_1232	A	PF735_RIF_AC_00_GE_2_31	0	0	0	22
10735	454_RIF_1241	A	PF735_RIF_AC_00_GE_2_45	0	0	0	22
10735	454_RIF_1250	A	PF735_RIF_AC_00_GE_2_36	0	0	0	5
10735	454_RIF_1255	A	PF735_RIF_AC_00_GE_2_40	0	0	0	13
10735	454_RIF_1260	A	PF735_RIF_AC_00_GE_2_44	0	0	0	16
10735	454_RIF_1270	A	PF735_RIF_AC_00_MR_10	5	0	0	0
10735	454_RIF_1274	A	PF735_RIF_AC_00_GE_2_42	0	0	0	43
10735	454_RIF_1275	A	PF735_RIF_AC_00_GE_2_48	0	0	0	14

Isolate	Cluster	A/B	Gene ID	R	LRET	ET	G
10735	454_RIF_1277	A	PF735_RIF_AC_00_GE_2_58	0	0	0	44
10735	454_RIF_1295	A	PF735_RIF_AC_00_GE_2_63	0	0	0	26
10735	454_RIF_1297	A	PF735_RIF_AC_00_GE_2_57	0	0	0	13
10735	454_RIF_1298	A	PF735_RIF_AC_00_GE_2_64	0	0	0	6
10735	454_RIF_1304	A	PF735_RIF_AC_00_GE_2_60	0	0	0	25
10735	454_RIF_1309	A	PF735_RIF_AC_00_GE_2_67	0	0	0	9
10735	454_RIF_1318	A	PF735_RIF_AC_00_GE_2_71	0	0	0	12
10735	454_RIF_1329	A	PF735_RIF_AC_00_GE_2_65	0	0	0	11
10735	454_RIF_1340	A	PF735_RIF_AC_00_GE_2_68	0	0	0	12
10735	454_RIF_1343	A	PF735_RIF_AC_00_GE_2_80	0	0	0	31
10735	454_RIF_1352	A	PF735_RIF_AC_00_GE_2_69	0	0	0	38
10735	454_RIF_1392	A	PF735_RIF_AC_00_GE_2_87	0	0	0	5
10735	454_RIF_1427	A	PF735_RIF_AC_00_GE_2_91	0	0	0	30
10735	454_RIF_1434	A	PF735_RIF_AC_00_GE_2_95	0	0	0	34
10735	454_RIF_1442	A	PF735_RIF_AC_00_GE_2_94	0	0	0	18
10735	454_RIF_1473	A	PF735_RIF_AC_00_GE_2_98	0	0	0	17
10735	454_RIF_1474	A	PF735_RIF_AC_00_GE_2_97	0	0	0	5
10735	454_RIF_1475	A	PF735_RIF_AC_00_GE_2_100	0	0	0	14
10735	454_RIF_1694	A	PF735_RIF_AC_30_ET_29	0	0	3	0
10735	454_RIF_1779	A	PF735_RIF_AC_00_MR_27	3	0	0	0
10735	454_RIF_1804	A	PF735_RIF_AC_00_MR_31	9	0	0	0
10735	454_RIF_1822	A	PF735_RIF_AC_00_GE_2_112	0	0	0	4
10735	454_RIF_1846	A	PF735_RIF_AC_00_MR_35	4	0	0	0
10735	454_RIF_1887	A	PF735_RIF_AC_00_MR_36	11	0	0	0
10735	454_RIF_1898	A	PF735_RIF_AC_30_ET_31	0	0	4	0
10735	454_RIF_1457	AX	PF735_RIF_AC_00_GE_2_99	0	0	0	8
10735	454_RIF_1476	AX	PF735_RIF_AC_00_GE_2_101	0	0	7	5
10735	454_RIF_1554	AX	PF735_RIF_AC_00_GE_2_103	0	0	0	6
10735	454_RIF_1624	AX	PF735_RIF_AC_00_GE_2_108	0	0	0	4
10735	454_RIF_1627	AX	PF735_RIF_AC_00_GE_2_107	0	0	0	6
10735	454_RIF_1831	AX	PF735_RIF_AC_00_MR_34	3	0	0	0
10735	454_RIF_0006	B	PF735_RIF_AC_00_GE_2_29	0	0	0	41
10735	454_RIF_0069	B	PF735_RIF_AC_00_GE_2_20	0	0	0	31
10735	454_RIF_0133	B	PF735_RIF_AC_00_GE_2_75	0	0	0	25
10735	454_RIF_0239	B	PF735_RIF_AC_00_GE_2_86	0	0	0	31
10735	454_RIF_0271	B	PF735_RIF_AC_00_MR_5	6	0	0	15
10735	454_RIF_0518	B	PF735_RIF_AC_00_MR_3	14	0	0	20
10735	454_RIF_0672	B	PF735_RIF_AC_00_GE_2_77	0	0	0	36
10735	454_RIF_0680	B	PF735_RIF_AC_00_GE_2_90	0	0	0	15
10735	454_RIF_0755	B	PF735_RIF_AC_00_GE_2_49	15	0	8	20
10735	454_RIF_0803	B	PF735_RIF_AC_00_MR_11	7	0	10	105
10735	454_RIF_0906	B	PF735_RIF_AC_00_GE_2_23	2	0	5	31
10735	454_RIF_0940	B	PF735_RIF_AC_00_GE_2_53	8	0	0	79
10735	454_RIF_0985	B	PF735_RIF_AC_00_MR_12	9	0	0	69
10735	454_RIF_0993	B	PF735_RIF_AC_00_MR_13	17	0	0	86
10735	454_RIF_1231	B	PF735_RIF_AC_00_GE_2_28	0	0	0	23
10735	454_RIF_1249	B	PF735_RIF_AC_00_GE_2_34	0	0	0	17
10735	454_RIF_1313	B	PF735_RIF_AC_00_GE_2_56	0	0	0	59
10735	454_RIF_1319	B	PF735_RIF_AC_00_GE_2_62	0	0	0	12
10735	454_RIF_1345	B	PF735_RIF_AC_00_GE_2_73	0	0	0	55
10735	454_RIF_1348	B	PF735_RIF_AC_00_GE_2_72	0	0	0	29
10735	454_RIF_1349	B	PF735_RIF_AC_00_GE_2_83	0	0	0	28
10735	454_RIF_1356	B	PF735_RIF_AC_00_GE_2_76	0	0	0	19
10735	454_RIF_1364	B	PF735_RIF_AC_00_GE_2_70	0	0	0	13
10735	454_RIF_1365	B	PF735_RIF_AC_00_GE_2_79	0	0	0	10
10735	454_RIF_1367	B	PF735_RIF_AC_00_GE_2_74	0	0	0	31
10735	454_RIF_1368	B	PF735_RIF_AC_00_GE_2_78	0	0	0	12
10735	454_RIF_1391	B	PF735_RIF_AC_00_GE_2_84	0	0	0	9
10735	454_RIF_1403	B	PF735_RIF_AC_00_GE_2_81	0	0	0	11
10735	454_RIF_1426	B	PF735_RIF_AC_00_GE_2_89	0	0	0	28
10735	454_RIF_1553	B	PF735_RIF_AC_00_MR_19	10	0	0	0
10735	454_RIF_1054	BX	PF735_RIF_AC_00_GE_2_105	0	0	0	7
10735	454_RIF_1765	BX	PF735_RIF_AC_00_MR_28	4	0	0	0

Isolate	Cluster	A/B	Gene ID	R	T	G
10739	454_RIF_0002	A	PF739_RIF_AC_00_GE_2_49	0	0	40
10739	454_RIF_0057	A	PF739_RIF_AC_00_GE_2_78	0	0	109
10739	454_RIF_0084	A	PF739_RIF_AC_00_GE_2_106	0	0	22
10739	454_RIF_0398	A	PF739_RIF_AC_00_GE_2_76	0	0	111
10739	454_RIF_0403	A	PF739_RIF_AC_00_GE_2_83	0	0	29
10739	454_RIF_0554	A	PF739_RIF_AC_00_GE_2_84	0	0	110
10739	454_RIF_0589	A	PF739_RIF_AC_00_GE_2_11	10	6	19
10739	454_RIF_0604	A	PF739_RIF_AC_00_GE_2_112	0	0	32
10739	454_RIF_0649	A	PF739_RIF_AC_00_GE_2_75	0	0	48
10739	454_RIF_0692	A	PF739_RIF_AC_00_GE_2_6	0	0	31
10739	454_RIF_0705	A	PF739_RIF_AC_00_GE_2_15	0	12	26
10739	454_RIF_0733	A	PF739_RIF_AC_40_TR_2	0	18.5	26
10739	454_RIF_0768	A	PF739_RIF_AC_00_MR_1	17	18	122
10739	454_RIF_0835	A	PF739_RIF_AC_00_GE_2_7	0	0	7
10739	454_RIF_0842	A	PF739_RIF_AC_00_GE_2_22	0	0	16
10739	454_RIF_0850	A	PF739_RIF_AC_00_GE_2_8	0	11	14
10739	454_RIF_0865	A	PF739_RIF_AC_00_GE_2_30	3	0	17
10739	454_RIF_0868	A	PF739_RIF_AC_00_GE_2_28	0	0	25
10739	454_RIF_0872	A	PF739_RIF_AC_00_GE_2_29	0	0	22
10739	454_RIF_0913	A	PF739_RIF_AC_40_TR_3	0	8	37
10739	454_RIF_0935	A	PF739_RIF_AC_40_TR_4	0	17	58
10739	454_RIF_0951	A	PF739_RIF_AC_00_GE_2_71	0	15	36
10739	454_RIF_0953	A	PF739_RIF_AC_00_GE_2_81	0	0	14
10739	454_RIF_1013	A	PF739_RIF_AC_00_GE_2_114	0	0	76
10739	454_RIF_1048	A	PF739_RIF_AC_00_GE_2_126	0	0	21
10739	454_RIF_1083	A	PF739_RIF_AC_00_GE_2_1	0	0	18
10739	454_RIF_1085	A	PF739_RIF_AC_00_GE_2_5	0	0	41
10739	454_RIF_1094	A	PF739_RIF_AC_00_GE_2_2	0	0	37
10739	454_RIF_1095	A	PF739_RIF_AC_00_GE_2_3	0	0	18
10739	454_RIF_1101	A	PF739_RIF_AC_00_GE_2_20	0	0	33
10739	454_RIF_1102	A	PF739_RIF_AC_00_GE_2_4	0	0	18
10739	454_RIF_1104	A	PF739_RIF_AC_00_GE_2_19	0	0	17
10739	454_RIF_1105	A	PF739_RIF_AC_00_GE_2_18	0	0	34
10739	454_RIF_1112	A	PF739_RIF_AC_00_GE_2_9	0	0	18
10739	454_RIF_1114	A	PF739_RIF_AC_00_GE_2_13	0	0	28
10739	454_RIF_1115	A	PF739_RIF_AC_00_GE_2_16	0	0	35
10739	454_RIF_1116	A	PF739_RIF_AC_00_GE_2_14	0	0	25
10739	454_RIF_1122	A	PF739_RIF_AC_00_GE_2_17	0	0	16
10739	454_RIF_1123	A	PF739_RIF_AC_00_GE_2_25	0	0	11
10739	454_RIF_1129	A	PF739_RIF_AC_00_GE_2_21	0	0	21
10739	454_RIF_1132	A	PF739_RIF_AC_00_GE_2_10	0	0	8
10739	454_RIF_1134	A	PF739_RIF_AC_00_GE_2_12	0	0	16
10739	454_RIF_1135	A	PF739_RIF_AC_00_GE_2_23	0	0	44
10739	454_RIF_1136	A	PF739_RIF_AC_00_GE_2_26	0	0	24
10739	454_RIF_1137	A	PF739_RIF_AC_00_GE_2_32	0	0	14
10739	454_RIF_1144	A	PF739_RIF_AC_00_GE_2_24	0	0	32
10739	454_RIF_1147	A	PF739_RIF_AC_00_GE_2_31	0	0	19
10739	454_RIF_1148	A	PF739_RIF_AC_00_GE_2_34	0	0	22
10739	454_RIF_1155	A	PF739_RIF_AC_00_GE_2_35	0	0	37
10739	454_RIF_1157	A	PF739_RIF_AC_00_GE_2_27	0	0	57
10739	454_RIF_1164	A	PF739_RIF_AC_00_GE_2_33	0	0	44
10739	454_RIF_1169	A	PF739_RIF_AC_00_GE_2_37	0	0	20
10739	454_RIF_1176	A	PF739_RIF_AC_00_GE_2_44	0	0	15
10739	454_RIF_1185	A	PF739_RIF_AC_00_GE_2_41	0	0	27
10739	454_RIF_1192	A	PF739_RIF_AC_00_GE_2_46	0	0	46
10739	454_RIF_1198	A	PF739_RIF_AC_00_GE_2_56	0	0	25
10739	454_RIF_1200	A	PF739_RIF_AC_00_GE_2_50	0	0	15
10739	454_RIF_1201	A	PF739_RIF_AC_00_GE_2_45	0	0	37
10739	454_RIF_1205	A	PF739_RIF_AC_00_GE_2_51	0	0	23
10739	454_RIF_1212	A	PF739_RIF_AC_00_GE_2_61	0	0	51
10739	454_RIF_1214	A	PF739_RIF_AC_00_GE_2_62	0	0	63
10739	454_RIF_1218	A	PF739_RIF_AC_00_GE_2_63	0	0	12
10739	454_RIF_1225	A	PF739_RIF_AC_00_GE_2_47	0	0	36
10739	454_RIF_1226	A	PF739_RIF_AC_00_GE_2_57	0	0	14
10739	454_RIF_1227	A	PF739_RIF_AC_00_GE_2_65	0	0	50
10739	454_RIF_1237	A	PF739_RIF_AC_00_GE_2_58	0	0	68
10739	454_RIF_1247	A	PF739_RIF_AC_00_GE_2_72	0	0	53
10739	454_RIF_1261	A	PF739_RIF_AC_00_GE_2_66	0	0	36

Isolate	Cluster	A/B	Gene ID	R	T	G
10739	454_RIF_1264	A	PF739_RIF_AC_00_GE_2_68	0	0	34
10739	454_RIF_1283	A	PF739_RIF_AC_00_GE_2_74	0	0	40
10739	454_RIF_1286	A	PF739_RIF_AC_00_GE_2_73	0	0	6
10739	454_RIF_1287	A	PF739_RIF_AC_00_GE_2_80	0	0	31
10739	454_RIF_1288	A	PF739_RIF_AC_00_GE_2_70	0	0	7
10739	454_RIF_1299	A	PF739_RIF_AC_00_GE_2_85	0	0	20
10739	454_RIF_1317	A	PF739_RIF_AC_00_GE_2_82	0	0	24
10739	454_RIF_1336	A	PF739_RIF_AC_00_GE_2_99	0	0	26
10739	454_RIF_1389	A	PF739_RIF_AC_40_TR_6	0	5	0
10739	454_RIF_1422	A	PF739_RIF_AC_00_GE_2_113	0	0	6
10739	454_RIF_1429	A	PF739_RIF_AC_00_GE_2_115	0	0	7
10739	454_RIF_1460	A	PF739_RIF_AC_00_GE_2_117	0	0	6
10739	454_RIF_1478	A	PF739_RIF_AC_00_GE_2_118	0	0	94
10739	454_RIF_1488	A	PF739_RIF_AC_00_GE_2_121	0	0	8
10739	454_RIF_1515	A	PF739_RIF_AC_40_TR_9	0	15	0
10739	454_RIF_1596	A	PF739_RIF_AC_00_GE_2_125	0	0	32
10739	454_RIF_1597	A	PF739_RIF_AC_50_MT_5	0	9	0
10739	454_RIF_1607	A	PF739_RIF_AC_00_GE_2_124	0	0	9
10739	454_RIF_1913	A	PF739_RIF_AC_12_PR_4	0	4	0
10739	454_RIF_1791	AX	PF739_RIF_AC_00_GE_2_127	0	0	2
10739	454_RIF_1897	AX	PF739_RIF_AC_00_GE_2_128	0	0	5
10739	454_RIF_1943	AX	PF739_RIF_AC_10_RN_4	8	0	0
10739	454_RIF_1944	AX	PF739_RIF_AC_40_TR_13	0	9	0
10739	454_RIF_0035	B	PF739_RIF_AC_00_GE_2_87	0	0	121
10739	454_RIF_0097	B	PF739_RIF_AC_00_GE_2_120	0	0	69
10739	454_RIF_0239	B	PF739_RIF_AC_00_GE_2_100	0	0	27
10739	454_RIF_0334	B	PF739_RIF_AC_00_GE_2_67	0	0	64
10739	454_RIF_0419	B	PF739_RIF_AC_00_GE_2_101	0	0	102
10739	454_RIF_0449	B	PF739_RIF_AC_00_GE_2_40	0	0	17
10739	454_RIF_0671	B	PF739_RIF_AC_00_GE_2_93	0	0	54
10739	454_RIF_0736	B	PF739_RIF_AC_00_GE_2_52	0	0	17
10739	454_RIF_0742	B	PF739_RIF_AC_00_GE_2_53	0	0	36
10739	454_RIF_0793	B	PF739_RIF_AC_00_GE_2_105	0	0	12
10739	454_RIF_0804	B	PF739_RIF_AC_00_GE_2_94	0	0	52
10739	454_RIF_0814	B	PF739_RIF_AC_00_GE_2_111	11	0	103
10739	454_RIF_0894	B	PF739_RIF_AC_00_GE_2_36	0	0	49
10739	454_RIF_0895	B	PF739_RIF_AC_00_GE_2_39	0	0	39
10739	454_RIF_0957	B	PF739_RIF_AC_00_GE_2_77	0	0	56
10739	454_RIF_0989	B	PF739_RIF_AC_00_GE_2_90	0	0	82
10739	454_RIF_0990	B	PF739_RIF_AC_00_GE_2_91	0	0	22
10739	454_RIF_0991	B	PF739_RIF_AC_00_GE_2_88	5	0	63
10739	454_RIF_0996	B	PF739_RIF_AC_00_GE_2_96	0	0	73
10739	454_RIF_1172	B	PF739_RIF_AC_00_GE_2_38	0	0	45
10739	454_RIF_1183	B	PF739_RIF_AC_00_GE_2_42	0	0	36
10739	454_RIF_1188	B	PF739_RIF_AC_00_GE_2_43	0	0	6
10739	454_RIF_1210	B	PF739_RIF_AC_00_GE_2_48	0	0	17
10739	454_RIF_1220	B	PF739_RIF_AC_00_GE_2_54	0	0	96
10739	454_RIF_1221	B	PF739_RIF_AC_00_GE_2_55	0	0	42
10739	454_RIF_1234	B	PF739_RIF_AC_00_GE_2_59	0	0	46
10739	454_RIF_1325	B	PF739_RIF_AC_00_GE_2_86	0	0	66
10739	454_RIF_1332	B	PF739_RIF_AC_00_GE_2_95	0	0	12
10739	454_RIF_1350	B	PF739_RIF_AC_00_GE_2_97	0	0	20
10739	454_RIF_1357	B	PF739_RIF_AC_00_GE_2_89	0	0	13
10739	454_RIF_1371	B	PF739_RIF_AC_00_GE_2_102	0	9	88
10739	454_RIF_1372	B	PF739_RIF_AC_00_GE_2_92	0	0	54
10739	454_RIF_1374	B	PF739_RIF_AC_00_GE_2_98	0	0	34
10739	454_RIF_1382	B	PF739_RIF_AC_00_GE_2_104	0	0	35
10739	454_RIF_1385	B	PF739_RIF_AC_00_GE_2_103	0	0	48
10739	454_RIF_1396	B	PF739_RIF_AC_00_GE_2_107	0	0	57
10739	454_RIF_1397	B	PF739_RIF_AC_00_GE_2_108	0	0	104
10739	454_RIF_1408	B	PF739_RIF_AC_00_GE_2_109	0	0	129
10739	454_RIF_1415	B	PF739_RIF_AC_00_GE_2_110	0	0	22
10739	454_RIF_1449	B	PF739_RIF_AC_00_GE_2_116	0	0	11
10739	454_RIF_1489	B	PF739_RIF_AC_00_GE_2_119	0	0	9
10739	454_RIF_1540	B	PF739_RIF_AC_00_GE_2_122	0	0	69
10739	454_RIF_1546	B	PF739_RIF_AC_00_GE_2_123	0	0	70
10739	454_RIF_1663	B	PF739_RIF_AC_50_MT_7	0	5	0
10739	454_RIF_1581	BX	PF739_RIF_AC_00_MR_6	2	0	0
10739	454_RIF_1933	BX	PF739_RIF_AC_00_GE_2_129	0	0	133

Isolate	Cluster	A/B	Gene ID	R	ET	ND-60 (S?)	G
10747	454_RIF_0001	A	PF747_RIF_AC_40_ET_2_23	14	21	17	0
10747	454_RIF_0002	A	PF747_RIF_AC_40_ET_2_69	5	31	22	0
10747	454_RIF_0003	A	PF747_RIF_AC_60_ND_2_4	38	45	31	0
10747	454_RIF_0023	A	PF747_RIF_AC_60_ND_2_24	0	0	14	0
10747	454_RIF_0027	A	PF747_RIF_AC_40_ET_2_73	0	17	9	0
10747	454_RIF_0038	A	PF747_RIF_AC_40_ET_2_100	0	8	12	0
10747	454_RIF_0041	A	PF747_RIF_AC_40_ET_2_125	0	19	0	0
10747	454_RIF_0123	A	PF747_RIF_AC_20_RN_2_33	11	0	0	0
10747	454_RIF_0213	A	PF747_RIF_AC_20_RN_2_107	6	0	0	0
10747	454_RIF_0232	A	PF747_RIF_AC_60_ND_2_96	0	0	17	0
10747	454_RIF_0255	A	PF747_RIF_AC_20_RN_2_81	9	0	0	0
10747	454_RIF_0266	A	PF747_RIF_AC_60_ND_2_10	0	7	8	0
10747	454_RIF_0280	A	PF747_RIF_AC_20_RN_2_31	32	18	0	0
10747	454_RIF_0297	A	PF747_RIF_AC_20_RN_2_114	21	0	0	0
10747	454_RIF_0372	A	PF747_RIF_AC_40_ET_2_71	0	3	0	0
10747	454_RIF_0374	A	PF747_RIF_AC_20_RN_2_10	11	0	0	0
10747	454_RIF_0385	A	PF747_RIF_AC_60_ND_2_100	0	0	10	0
10747	454_RIF_0441	A	PF747_RIF_AC_40_ET_2_10	37	14	6	5
10747	454_RIF_0479	A	PF747_RIF_AC_40_ET_2_37	206	90	96	4
10747	454_RIF_0492	A	PF747_RIF_AC_00_GE_35	0	0	0	6
10747	454_RIF_0526	A	PF747_RIF_AC_00_GE_3	0	3	15	16
10747	454_RIF_0536	A	PF747_RIF_AC_20_RN_2_68	9	0	0	0
10747	454_RIF_0551	A	PF747_RIF_AC_20_RN_2_63	2	0	0	0
10747	454_RIF_0577	A	PF747_RIF_AC_40_ET_2_4	16	13	14	0
10747	454_RIF_0596	A	PF747_RIF_AC_40_ET_2_6	6	15	7	0
10747	454_RIF_0598	A	PF747_RIF_AC_60_ND_2_5	11	29	23	0
10747	454_RIF_0622	A	PF747_RIF_AC_60_ND_2_14	79	23	26	0
10747	454_RIF_0641	A	PF747_RIF_AC_40_ET_2_36	182	46	21	0
10747	454_RIF_0650	A	PF747_RIF_AC_20_RN_2_62	13	0	9	7
10747	454_RIF_0683	A	PF747_RIF_AC_40_ET_2_1	28	22	16	0
10747	454_RIF_0699	A	PF747_RIF_AC_20_RN_2_2	17	32	0	0
10747	454_RIF_0701	A	PF747_RIF_AC_20_RN_2_3	17	7	4	0
10747	454_RIF_0707	A	PF747_RIF_AC_40_ET_2_11	72	50	19	0
10747	454_RIF_0714	A	PF747_RIF_AC_20_RN_2_7	284	100	53	0
10747	454_RIF_0731	A	PF747_RIF_AC_40_ET_2_22	15	13	10	0
10747	454_RIF_0740	A	PF747_RIF_AC_40_ET_2_25	72	25	12	0
10747	454_RIF_0743	A	PF747_RIF_AC_40_ET_2_18	58	17	14	0
10747	454_RIF_0746	A	PF747_RIF_AC_60_ND_2_15	19	9	9	0
10747	454_RIF_0749	A	PF747_RIF_AC_40_ET_2_19	108	25	13	0
10747	454_RIF_0756	A	PF747_RIF_AC_40_ET_2_26	94	30	11	0
10747	454_RIF_0767	A	PF747_RIF_AC_40_ET_2_35	84	23	14	0
10747	454_RIF_0773	A	PF747_RIF_AC_20_RN_2_32	26	26	11	0
10747	454_RIF_0774	A	PF747_RIF_AC_40_ET_2_40	145	40	20	0
10747	454_RIF_0775	A	PF747_RIF_AC_40_ET_2_41	20	23	11	0
10747	454_RIF_0781	A	PF747_RIF_AC_20_RN_2_35	13	15	0	8
10747	454_RIF_0783	A	PF747_RIF_AC_20_RN_2_37	8	22	0	0
10747	454_RIF_0832	A	PF747_RIF_AC_40_ET_2_2	5	8	0	0
10747	454_RIF_0836	A	PF747_RIF_AC_20_RN_2_1	32	0	3	0
10747	454_RIF_0851	A	PF747_RIF_AC_40_ET_2_3	0	9	11	0
10747	454_RIF_0862	A	PF747_RIF_AC_20_RN_2_9	7	6	0	0
10747	454_RIF_0869	A	PF747_RIF_AC_20_RN_2_5	24	7	0	0
10747	454_RIF_0902	A	PF747_RIF_AC_20_RN_2_16	27	6	0	0
10747	454_RIF_0903	A	PF747_RIF_AC_40_ET_2_15	0	10	0	5
10747	454_RIF_0914	A	PF747_RIF_AC_60_ND_2_12	6	0	10	0
10747	454_RIF_0916	A	PF747_RIF_AC_40_ET_2_24	20	10	0	0
10747	454_RIF_0923	A	PF747_RIF_AC_20_RN_2_22	26	4	0	0
10747	454_RIF_0932	A	PF747_RIF_AC_40_ET_2_30	0	5	0	4
10747	454_RIF_0936	A	PF747_RIF_AC_20_RN_2_28	30	14	0	0
10747	454_RIF_0942	A	PF747_RIF_AC_00_GE_7	0	0	6	14

Isolate	Cluster	A/B	Gene ID	R	ET	ND-60 (S?)	G
10747	454_RIF_0949	A	PF747_RIF_AC_40_ET_2_72	0	11	0	0
10747	454_RIF_0955	A	PF747_RIF_AC_40_ET_2_32	0	11	3	0
10747	454_RIF_0956	A	PF747_RIF_AC_40_ET_2_138	0	15	0	0
10747	454_RIF_0980	A	PF747_RIF_AC_20_RN_2_41	14	0	4	0
10747	454_RIF_0981	A	PF747_RIF_AC_40_ET_2_43	0	24	23	0
10747	454_RIF_0997	A	PF747_RIF_AC_00_GE_10	8	0	0	10
10747	454_RIF_1023	A	PF747_RIF_AC_60_ND_2_46	0	25	12	0
10747	454_RIF_1027	A	PF747_RIF_AC_40_ET_2_65	0	10	9	0
10747	454_RIF_1070	A	PF747_RIF_AC_60_ND_2_61	16	0	10	0
10747	454_RIF_1073	A	PF747_RIF_AC_60_ND_2_78	21	0	7	0
10747	454_RIF_1076	A	PF747_RIF_AC_40_ET_2_104	0	2	2	0
10747	454_RIF_1103	A	PF747_RIF_AC_60_ND_2_1	0	0	5	0
10747	454_RIF_1118	A	PF747_RIF_AC_20_RN_2_4	7	0	0	0
10747	454_RIF_1139	A	PF747_RIF_AC_40_ET_2_7	0	11	0	0
10747	454_RIF_1152	A	PF747_RIF_AC_00_GE_1	0	0	0	13
10747	454_RIF_1162	A	PF747_RIF_AC_20_RN_2_8	9	0	0	0
10747	454_RIF_1173	A	PF747_RIF_AC_20_RN_2_14	7	0	0	0
10747	454_RIF_1177	A	PF747_RIF_AC_20_RN_2_13	8	0	0	0
10747	454_RIF_1217	A	PF747_RIF_AC_20_RN_2_18	30	0	0	0
10747	454_RIF_1222	A	PF747_RIF_AC_60_ND_2_20	0	0	31	0
10747	454_RIF_1228	A	PF747_RIF_AC_40_ET_2_17	0	6	0	0
10747	454_RIF_1284	A	PF747_RIF_AC_40_ET_2_34	0	7	0	0
10747	454_RIF_1285	A	PF747_RIF_AC_40_ET_2_39	0	5	0	0
10747	454_RIF_1291	A	PF747_RIF_AC_60_ND_2_25	0	0	10	0
10747	454_RIF_1296	A	PF747_RIF_AC_20_RN_3	6	0	0	0
10747	454_RIF_1300	A	PF747_RIF_AC_20_RN_2_36	34	0	0	0
10747	454_RIF_1310	A	PF747_RIF_AC_40_ET_2_38	0	6	0	0
10747	454_RIF_1324	A	PF747_RIF_AC_20_RN_2_40	7	0	0	0
10747	454_RIF_1358	A	PF747_RIF_AC_20_RN_2_43	13	0	0	0
10747	454_RIF_1359	A	PF747_RIF_AC_20_RN_2_50	6	0	0	0
10747	454_RIF_1398	A	PF747_RIF_AC_20_RN_2_49	11	0	0	0
10747	454_RIF_1409	A	PF747_RIF_AC_20_RN_2_52	9	0	0	0
10747	454_RIF_1413	A	PF747_RIF_AC_60_ND_2_40	0	0	11	0
10747	454_RIF_1419	A	PF747_RIF_AC_20_RN_2_53	17	0	0	4
10747	454_RIF_1432	A	PF747_RIF_AC_40_ET_2_57	0	15	0	0
10747	454_RIF_1439	A	PF747_RIF_AC_40_ET_2_114	0	10	0	0
10747	454_RIF_1450	A	PF747_RIF_AC_40_ET_2_63	0	6	0	0
10747	454_RIF_1453	A	PF747_RIF_AC_20_RN_2_57	23	0	0	0
10747	454_RIF_1466	A	PF747_RIF_AC_20_RN_2_59	8	0	0	0
10747	454_RIF_1468	A	PF747_RIF_AC_60_ND_2_47	0	0	6	0
10747	454_RIF_1490	A	PF747_RIF_AC_40_ET_2_68	0	8	0	0
10747	454_RIF_1503	A	PF747_RIF_AC_40_ET_2_70	0	8	0	0
10747	454_RIF_1525	A	PF747_RIF_AC_40_ET_2_74	0	3	0	0
10747	454_RIF_1532	A	PF747_RIF_AC_20_RN_2_60	9	0	0	0
10747	454_RIF_1547	A	PF747_RIF_AC_40_ET_2_77	0	2	0	0
10747	454_RIF_1550	A	PF747_RIF_AC_20_RN_2_66	2	0	0	0
10747	454_RIF_1552	A	PF747_RIF_AC_40_ET_2_80	0	13	0	0
10747	454_RIF_1556	A	PF747_RIF_AC_40_ET_2_81	0	2	0	0
10747	454_RIF_1612	A	PF747_RIF_AC_40_ET_2_91	0	2	0	0
10747	454_RIF_1613	A	PF747_RIF_AC_40_ET_2_93	0	5	0	0
10747	454_RIF_1626	A	PF747_RIF_AC_00_GE_26	0	0	0	9
10747	454_RIF_1635	A	PF747_RIF_AC_00_GE_25	0	0	0	12
10747	454_RIF_1637	A	PF747_RIF_AC_20_RN_2_77	8	0	0	0
10747	454_RIF_1638	A	PF747_RIF_AC_20_RN_2_76	7	0	0	0
10747	454_RIF_1640	A	PF747_RIF_AC_60_ND_2_65	0	0	4	0
10747	454_RIF_1645	A	PF747_RIF_AC_60_ND_2_66	0	0	2	0
10747	454_RIF_1648	A	PF747_RIF_AC_00_GE_27	0	0	0	6
10747	454_RIF_1652	A	PF747_RIF_AC_60_ND_2_67	0	6	3	0
10747	454_RIF_1653	A	PF747_RIF_AC_00_GE_28	0	0	0	8

Isolate	Cluster	A/B	Gene ID	R	ET	ND-60 (S?)	G
10747	454_RIF_1655	A	PF747_RIF_AC_40_ET_2_101	0	12	0	0
10747	454_RIF_1666	A	PF747_RIF_AC_60_ND_2_72	0	0	8	0
10747	454_RIF_1673	A	PF747_RIF_AC_60_ND_2_73	0	0	6	0
10747	454_RIF_1701	A	PF747_RIF_AC_00_GE_30	0	0	0	5
10747	454_RIF_1718	A	PF747_RIF_AC_60_ND_2_82	0	0	2	0
10747	454_RIF_1724	A	PF747_RIF_AC_60_ND_2_80	0	0	9	0
10747	454_RIF_1739	A	PF747_RIF_AC_40_ET_2_106	0	5	0	0
10747	454_RIF_1740	A	PF747_RIF_AC_60_ND_2_81	0	0	9	0
10747	454_RIF_1744	A	PF747_RIF_AC_40_ET_2_107	0	3	0	0
10747	454_RIF_1746	A	PF747_RIF_AC_20_RN_8	6	0	0	0
10747	454_RIF_1747	A	PF747_RIF_AC_60_ND_2_85	0	0	3	0
10747	454_RIF_1760	A	PF747_RIF_AC_60_ND_2_88	0	0	2	0
10747	454_RIF_1780	A	PF747_RIF_AC_40_ET_2_113	0	5	0	0
10747	454_RIF_1794	A	PF747_RIF_AC_40_ET_2_117	0	27	0	0
10747	454_RIF_1797	A	PF747_RIF_AC_20_RN_2_93	64	0	0	0
10747	454_RIF_1848	A	PF747_RIF_AC_20_RN_2_96	11	9	0	0
10747	454_RIF_1856	A	PF747_RIF_AC_40_ET_2_128	0	4	0	0
10747	454_RIF_1859	A	PF747_RIF_AC_20_RN_2_101	3	9	0	0
10747	454_RIF_1866	A	PF747_RIF_AC_40_ET_2_131	0	3	0	0
10747	454_RIF_1867	A	PF747_RIF_AC_20_RN_2_100	8	2	0	0
10747	454_RIF_1871	A	PF747_RIF_AC_20_RN_2_104	3	0	0	0
10747	454_RIF_1903	A	PF747_RIF_AC_60_ND_2_106	0	0	12	0
10747	454_RIF_1939	A	PF747_RIF_AC_40_ET_2_141	0	0	0	0
10747	454_RIF_0823	AX	PF747_RIF_AC_20_RN_2_64	11	14	7	0
10747	454_RIF_0828	AX	PF747_RIF_AC_60_ND_2_90	0	10	12	0
10747	454_RIF_1036	AX	PF747_RIF_AC_20_RN_2_58	9	0	0	4
10747	454_RIF_1063	AX	PF747_RIF_AC_40_ET_2_90	3	4	0	0
10747	454_RIF_1454	AX	PF747_RIF_AC_20_RN_2_56	7	0	0	0
10747	454_RIF_1533	AX	PF747_RIF_AC_20_RN_2_61	8	0	13	0
10747	454_RIF_1562	AX	PF747_RIF_AC_20_RN_2_69	21	0	0	0
10747	454_RIF_1563	AX	PF747_RIF_AC_20_RN_2_65	6	0	0	0
10747	454_RIF_1594	AX	PF747_RIF_AC_00_GE_23	0	0	0	5
10747	454_RIF_1632	AX	PF747_RIF_AC_40_ET_2_97	0	3	0	0
10747	454_RIF_1636	AX	PF747_RIF_AC_20_RN_2_74	21	0	0	0
10747	454_RIF_1665	AX	PF747_RIF_AC_60_ND_2_71	0	0	2	0
10747	454_RIF_1668	AX	PF747_RIF_AC_20_RN_2_80	4	0	0	0
10747	454_RIF_1722	AX	PF747_RIF_AC_20_RN_2_88	4	0	0	0
10747	454_RIF_1754	AX	PF747_RIF_AC_20_RN_2_92	6	0	0	0
10747	454_RIF_1755	AX	PF747_RIF_AC_40_ET_2_108	0	2	0	0
10747	454_RIF_1799	AX	PF747_RIF_AC_60_ND_2_93	0	0	1	0
10747	454_RIF_1827	AX	PF747_RIF_AC_60_ND_2_94	0	0	2	0
10747	454_RIF_1837	AX	PF747_RIF_AC_40_ET_2_123	0	4	0	0
10747	454_RIF_1853	AX	PF747_RIF_AC_40_ET_2_127	0	3	0	0
10747	454_RIF_1881	AX	PF747_RIF_AC_60_ND_2_102	0	0	3	0
10747	454_RIF_1886	AX	PF747_RIF_AC_20_RN_2_102	5	0	0	0
10747	454_RIF_1888	AX	PF747_RIF_AC_60_ND_2_101	0	0	2	0
10747	454_RIF_1909	AX	PF747_RIF_AC_20_RN_2_106	4	0	0	0
10747	454_RIF_1915	AX	PF747_RIF_AC_60_ND_2_104_	0	0	8	0
10747	454_RIF_1921	AX	PF747_RIF_AC_20_RN_2_110	2	0	0	0
10747	454_RIF_1928	AX	PF747_RIF_AC_60_ND_2_109	0	0	4	0
10747	454_RIF_1951	AX	PF747_RIF_AC_00_GE_38	0	0	0	11
10747	454_RIF_0012	B	PF747_RIF_AC_60_ND_2_44	7	33	13	0
10747	454_RIF_0017	B	PF747_RIF_AC_40_ET_2_124	0	2	0	0
10747	454_RIF_0034	B	PF747_RIF_AC_60_ND_2_113	0	0	1	0
10747	454_RIF_0042	B	PF747_RIF_AC_00_GE_5	0	0	0	14
10747	454_RIF_0043	B	PF747_RIF_AC_60_ND_2_63	0	0	8	0
10747	454_RIF_0046	B	PF747_RIF_AC_40_ET_2_129	0	42	0	0
10747	454_RIF_0060	B	PF747_RIF_AC_40_ET_2_48	0	58	32	0
10747	454_RIF_0087	B	PF747_RIF_AC_40_ET_2_109	0	3	0	0

Isolate	Cluster	A/B	Gene ID	R	ET	ND-60 (S?)	G
10747	454_RIF_0203	B	PF747_RIF_AC_60_ND_2_38	0	0	13	0
10747	454_RIF_0239	B	PF747_RIF_AC_40_ET_2_67	0	7	0	8
10747	454_RIF_0240	B	PF747_RIF_AC_60_ND_2_37	0	14	8	0
10747	454_RIF_0287	B	PF747_RIF_AC_60_ND_2_107	0	0	18	0
10747	454_RIF_0289	B	PF747_RIF_AC_60_ND_2_60	0	25	9	0
10747	454_RIF_0317	B	PF747_RIF_AC_20_RN_2_15	11	41	43	0
10747	454_RIF_0380	B	PF747_RIF_AC_00_GE_2	0	6	0	10
10747	454_RIF_0412	B	PF747_RIF_AC_40_ET_2_137	0	4	0	0
10747	454_RIF_0419	B	PF747_RIF_AC_00_GE_13	0	47	35	36
10747	454_RIF_0559	B	PF747_RIF_AC_00_GE_9	21	23	17	22
10747	454_RIF_0561	B	PF747_RIF_AC_60_ND_2_32	0	8	7	0
10747	454_RIF_0565	B	PF747_RIF_AC_20_RN_2_44	13	15	32	10
10747	454_RIF_0605	B	PF747_RIF_AC_60_ND_2_8	32	24	27	0
10747	454_RIF_0638	B	PF747_RIF_AC_60_ND_2_21	7	37	34	6
10747	454_RIF_0661	B	PF747_RIF_AC_40_ET_2_47	0	11	0	0
10747	454_RIF_0662	B	PF747_RIF_AC_60_ND_2_30	24	5	13	0
10747	454_RIF_0669	B	PF747_RIF_AC_60_ND_2_35	18	54	62	2
10747	454_RIF_0672	B	PF747_RIF_AC_60_ND_2_34	0	23	23	8
10747	454_RIF_0678	B	PF747_RIF_AC_40_ET_2_53	3	7	7	8
10747	454_RIF_0747	B	PF747_RIF_AC_60_ND_2_17	0	21	27	9
10747	454_RIF_0759	B	PF747_RIF_AC_60_ND_2_19	6	10	12	0
10747	454_RIF_0784	B	PF747_RIF_AC_40_ET_2_42	6	7	8	0
10747	454_RIF_0790	B	PF747_RIF_AC_40_ET_2_44	6	6	17	0
10747	454_RIF_0796	B	PF747_RIF_AC_40_ET_2_139	0	1	0	0
10747	454_RIF_0798	B	PF747_RIF_AC_60_ND_2_36	7	38	24	0
10747	454_RIF_0799	B	PF747_RIF_AC_40_ET_2_52	5	29	28	0
10747	454_RIF_0891	B	PF747_RIF_AC_20_RN_2_83	11	0	0	0
10747	454_RIF_0893	B	PF747_RIF_AC_40_ET_2_13	0	14	23	0
10747	454_RIF_0920	B	PF747_RIF_AC_40_ET_2_21	0	8	2	0
10747	454_RIF_0987	B	PF747_RIF_AC_60_ND_2_31	0	11	23	0
10747	454_RIF_1008	B	PF747_RIF_AC_40_ET_2_112	0	3	0	0
10747	454_RIF_1046	B	PF747_RIF_AC_20_RN_5	7	12	0	0
10747	454_RIF_1047	B	PF747_RIF_AC_40_ET_2_75	0	17	6	0
10747	454_RIF_1059	B	PF747_RIF_AC_40_ET_2_88	0	4	4	0
10747	454_RIF_1254	B	PF747_RIF_AC_40_ET_2_28	0	9	0	0
10747	454_RIF_1306	B	PF747_RIF_AC_20_RN_2_39	5	0	0	0
10747	454_RIF_1375	B	PF747_RIF_AC_40_ET_2_45	0	9	0	0
10747	454_RIF_1376	B	PF747_RIF_AC_40_ET_2_50	0	6	0	0
10747	454_RIF_1510	B	PF747_RIF_AC_00_GE_18	0	0	0	3
10747	454_RIF_1559	B	PF747_RIF_AC_00_GE_21	0	0	0	5
10747	454_RIF_1586	B	PF747_RIF_AC_60_ND_2_57	0	0	4	0
10747	454_RIF_1608	B	PF747_RIF_AC_20_RN_2_71	3	0	0	0
10747	454_RIF_1631	B	PF747_RIF_AC_60_ND_2_62	0	0	5	0
10747	454_RIF_1672	B	PF747_RIF_AC_20_RN_2_82	5	0	0	0
10747	454_RIF_1676	B	PF747_RIF_AC_20_RN_2_78	7	0	0	0
10747	454_RIF_1679	B	PF747_RIF_AC_20_RN_2_85	3	0	0	0
10747	454_RIF_1707	B	PF747_RIF_AC_20_RN_2_87	6	0	0	0
10747	454_RIF_1750	B	PF747_RIF_AC_40_ET_2_105	0	9	0	0
10747	454_RIF_1781	B	PF747_RIF_AC_00_GE_33	0	0	0	2
10747	454_RIF_1833	B	PF747_RIF_AC_40_ET_2_122	0	1	0	0
10747	454_RIF_1844	B	PF747_RIF_AC_40_ET_2_126	0	3	0	0
10747	454_RIF_1041	BX	PF747_RIF_AC_60_ND_2_50	0	10	7	0
10747	454_RIF_1054	BX	PF747_RIF_AC_40_ET_2_84	0	6	0	0
10747	454_RIF_1544	BX	PF747_RIF_AC_40_ET_2_76	0	6	0	0
10747	454_RIF_1575	BX	PF747_RIF_AC_40_ET_2_87	0	7	0	0
10747	454_RIF_1582	BX	PF747_RIF_AC_40_ET_2_82	0	7	0	0
10747	454_RIF_1656	BX	PF747_RIF_AC_60_ND_2_69	0	0	10	0
10747	454_RIF_1839	BX	PF747_RIF_AC_40_ET_2_119	0	7	0	0
10747	454_RIF_1863	BX	PF747_RIF_AC_40_ET_2_130	0	3	0	0
10747	454_RIF_1884	BX	PF747_RIF_AC_00_GE_36	0	0	0	2
10747	454_RIF_1891	BX	PF747_RIF_AC_40_ET_2_135	0	0	0	0
10747	454_RIF_1950	BX	PF747_RIF_AC_60_ND_2_112	0	0	10	0
10747	454_RIF_1953	BX	PF747_RIF_AC_20_RN_2_112	7	0	0	0
10747	454_RIF_1892	XB	PF747_RIF_AC_40_ET_2_136	0	0	0	0

Isolate	Cluster	A/B	Gene ID	R	LT	S-ER2	G
10761	454_RIF_0001	A	PF761_RIF_AC_34_SZ_2_30	0	0	17	19
10761	454_RIF_0005	A	PF761_RIF_AC_34_SZ_2_25	0	0	16	8
10761	454_RIF_0008	A	PF761_RIF_AC_00_GE_2_48	0	0	0	12
10761	454_RIF_0021	A	PF761_RIF_AC_12_MR_7	8	0	0	0
10761	454_RIF_0023	A	PF761_RIF_AC_34_SZ_2_43	0	0	17	0
10761	454_RIF_0027	A	PF761_RIF_AC_00_GE_2_39	3	0	18	33
10761	454_RIF_0071	A	PF761_RIF_AC_00_GE_2_83	0	0	0	6
10761	454_RIF_0120	A	PF761_RIF_AC_00_GE_2_63	0	0	0	8
10761	454_RIF_0123	A	PF761_RIF_AC_00_GE_2_37	0	0	0	24
10761	454_RIF_0237	A	PF761_RIF_AC_00_GE_2_73	0	0	4	12
10761	454_RIF_0251	A	PF761_RIF_AC_00_GE_2_103	0	0	0	17
10761	454_RIF_0298	A	PF761_RIF_AC_34_SZ_2_7	74	9	44	0
10761	454_RIF_0331	A	PF761_RIF_AC_00_GE_2_31	0	0	0	6
10761	454_RIF_0339	A	PF761_RIF_AC_34_SZ_2_55	0	0	7	0
10761	454_RIF_0350	A	PF761_RIF_AC_34_SZ_2_54	0	0	13	0
10761	454_RIF_0386	A	PF761_RIF_AC_34_SZ_2_113	0	0	3	0
10761	454_RIF_0400	A	PF761_RIF_AC_34_SZ_2_42	0	0	10	0
10761	454_RIF_0403	A	PF761_RIF_AC_00_GE_2_68	0	0	0	13
10761	454_RIF_0405	A	PF761_RIF_AC_34_SZ_2_88	0	0	7	0
10761	454_RIF_0408	A	PF761_RIF_AC_34_SZ_2_126	0	0	4	0
10761	454_RIF_0451	A	PF761_RIF_AC_00_ER_2	47	0	977	7
10761	454_RIF_0476	A	PF761_RIF_AC_12_MR_4	14	0	29	23
10761	454_RIF_0494	A	PF761_RIF_AC_16_MR_1	19	0	8	5
10761	454_RIF_0498	A	PF761_RIF_AC_34_SZ_2_71	0	0	9	0
10761	454_RIF_0503	A	PF761_RIF_AC_00_GE_2_3	19	0	56	12
10761	454_RIF_0514	A	PF761_RIF_AC_00_GE_2_109	0	0	0	9
10761	454_RIF_0523	A	PF761_RIF_AC_16_MR_8	17	0	58	12
10761	454_RIF_0533	A	PF761_RIF_AC_16_MR_10	37	0	12	11
10761	454_RIF_0536	A	PF761_RIF_AC_34_SZ_2_67	7	0	18	0
10761	454_RIF_0550	A	PF761_RIF_AC_16_MR_13	18	0	66	14
10761	454_RIF_0551	A	PF761_RIF_AC_34_SZ_2_50	0	0	7	7
10761	454_RIF_0554	A	PF761_RIF_AC_12_MR_3	6	0	44	6
10761	454_RIF_0583	A	PF761_RIF_AC_34_SZ_2_4	0	0	7	0
10761	454_RIF_0597	A	PF761_RIF_AC_00_GE_2_11	99	0	73	19
10761	454_RIF_0604	A	PF761_RIF_AC_00_GE_2_13	0	0	30	23
10761	454_RIF_0608	A	PF761_RIF_AC_34_SZ_2_23	7	0	6	7
10761	454_RIF_0609	A	PF761_RIF_AC_34_SZ_2_24	0	0	10	0
10761	454_RIF_0610	A	PF761_RIF_AC_00_GE_2_19	0	0	9	23
10761	454_RIF_0612	A	PF761_RIF_AC_34_SZ_2_26	10	0	31	7
10761	454_RIF_0636	A	PF761_RIF_AC_00_GE_2_36	17	0	42	5
10761	454_RIF_0647	A	PF761_RIF_AC_00_GE_2_92	0	0	0	19
10761	454_RIF_0650	A	PF761_RIF_AC_00_GE_2_121	0	0	0	15
10761	454_RIF_0690	A	PF761_RIF_AC_16_MR_6	10	0	6	0
10761	454_RIF_0692	A	PF761_RIF_AC_34_SZ_2_6	0	0	9	3
10761	454_RIF_0694	A	PF761_RIF_AC_34_SZ_2_2	0	0	23	28
10761	454_RIF_0695	A	PF761_RIF_AC_16_MR_28	8	0	0	0
10761	454_RIF_0704	A	PF761_RIF_AC_16_MR_3	19	0	16	0
10761	454_RIF_0705	A	PF761_RIF_AC_34_SZ_2_31	0	0	4	0
10761	454_RIF_0706	A	PF761_RIF_AC_00_GE_2_9	0	0	16	11
10761	454_RIF_0708	A	PF761_RIF_AC_34_SZ_2_13	0	0	13	14
10761	454_RIF_0713	A	PF761_RIF_AC_34_SZ_2_17	0	0	11	5
10761	454_RIF_0716	A	PF761_RIF_AC_34_SZ_2_19	0	0	20	7
10761	454_RIF_0719	A	PF761_RIF_AC_00_GE_2_10	0	0	5	16
10761	454_RIF_0724	A	PF761_RIF_AC_00_GE_2_20	0	0	9	8
10761	454_RIF_0735	A	PF761_RIF_AC_34_SZ_2_84	0	0	5	0
10761	454_RIF_0745	A	PF761_RIF_AC_00_GE_2_113	0	0	0	4
10761	454_RIF_0757	A	PF761_RIF_AC_00_GE_2_34	0	0	30	26
10761	454_RIF_0762	A	PF761_RIF_AC_00_GE_2_35	9	0	22	6
10761	454_RIF_0776	A	PF761_RIF_AC_16_MR_14	11	0	0	6
10761	454_RIF_0789	A	PF761_RIF_AC_16_MR_18	40	0	11	0
10761	454_RIF_0816	A	PF761_RIF_AC_00_GE_2_85	0	0	24	64
10761	454_RIF_0819	A	PF761_RIF_AC_16_MR_22	23	0	33	0
10761	454_RIF_0831	A	PF761_RIF_AC_34_SZ_2_46	0	0	10	0
10761	454_RIF_0833	A	PF761_RIF_AC_34_SZ_2_3	0	0	11	0
10761	454_RIF_0842	A	PF761_RIF_AC_00_GE_2_42	0	0	0	5
10761	454_RIF_0863	A	PF761_RIF_AC_34_SZ_2_14	11	0	8	0
10761	454_RIF_0866	A	PF761_RIF_AC_00_GE_2_5	0	0	12	9
10761	454_RIF_0868	A	PF761_RIF_AC_00_GE_2_79	0	0	0	6
10761	454_RIF_0873	A	PF761_RIF_AC_00_GE_2_4	0	0	0	18

Isolate	Cluster	A/B	Gene ID	R	LT	S-ER2	G
10761	454_RIF_0876	A	PF761_RIF_AC_34_SZ_2_10	0	0	9	0
10761	454_RIF_0877	A	PF761_RIF_AC_34_SZ_2_12	0	0	11	4
10761	454_RIF_0886	A	PF761_RIF_AC_00_GE_2_77	0	0	0	9
10761	454_RIF_0887	A	PF761_RIF_AC_34_SZ_2_34	0	0	17	0
10761	454_RIF_0892	A	PF761_RIF_AC_34_SZ_2_21	0	0	8	0
10761	454_RIF_0901	A	PF761_RIF_AC_00_GE_2_17	0	0	0	15
10761	454_RIF_0904	A	PF761_RIF_AC_00_GE_2_18	0	0	5	10
10761	454_RIF_0909	A	PF761_RIF_AC_00_GE_2_21	0	0	0	14
10761	454_RIF_0915	A	PF761_RIF_AC_00_GE_2_23	0	0	2	6
10761	454_RIF_0921	A	PF761_RIF_AC_00_GE_2_24	0	0	3	15
10761	454_RIF_0927	A	PF761_RIF_AC_00_GE_2_89	0	0	0	8
10761	454_RIF_0931	A	PF761_RIF_AC_16_MR_12	5	0	0	3
10761	454_RIF_0944	A	PF761_RIF_AC_00_GE_2_32	0	0	0	19
10761	454_RIF_0949	A	PF761_RIF_AC_34_SZ_2_40	0	0	22	0
10761	454_RIF_0953	A	PF761_RIF_AC_34_SZ_2_63	3	0	5	0
10761	454_RIF_1001	A	PF761_RIF_AC_34_SZ_2_69	0	0	11	0
10761	454_RIF_1013	A	PF761_RIF_AC_00_GE_2_110	0	0	0	10
10761	454_RIF_1062	A	PF761_RIF_AC_34_SZ_2_127	0	0	6	0
10761	454_RIF_1087	A	PF761_RIF_AC_34_SZ_2_1	0	0	15	0
10761	454_RIF_1092	A	PF761_RIF_AC_00_GE_2_1	0	0	0	9
10761	454_RIF_1108	A	PF761_RIF_AC_00_GE_2_7	0	0	0	11
10761	454_RIF_1109	A	PF761_RIF_AC_34_SZ_2_8	0	0	11	0
10761	454_RIF_1119	A	PF761_RIF_AC_00_GE_2_2	0	0	0	11
10761	454_RIF_1124	A	PF761_RIF_AC_34_SZ_2_9	0	0	6	0
10761	454_RIF_1145	A	PF761_RIF_AC_00_GE_2_12	0	0	0	11
10761	454_RIF_1154	A	PF761_RIF_AC_00_GE_2_6	0	0	0	22
10761	454_RIF_1159	A	PF761_RIF_AC_34_SZ_2_20	0	0	18	0
10761	454_RIF_1160	A	PF761_RIF_AC_00_GE_2_8	0	0	0	13
10761	454_RIF_1181	A	PF761_RIF_AC_34_SZ_2_28	0	0	7	0
10761	454_RIF_1262	A	PF761_RIF_AC_00_GE_2_29	0	0	0	4
10761	454_RIF_1278	A	PF761_RIF_AC_34_SZ_2_37	0	0	8	0
10761	454_RIF_1292	A	PF761_RIF_AC_16_MR_15	5	0	0	0
10761	454_RIF_1320	A	PF761_RIF_AC_00_GE_2_41	0	0	0	4
10761	454_RIF_1330	A	PF761_RIF_AC_34_SZ_2_51	0	0	5	0
10761	454_RIF_1366	A	PF761_RIF_AC_00_GE_2_52	0	0	0	34
10761	454_RIF_1420	A	PF761_RIF_AC_00_GE_2_66	0	0	0	15
10761	454_RIF_1424	A	PF761_RIF_AC_00_GE_2_61	0	0	0	17
10761	454_RIF_1472	A	PF761_RIF_AC_00_GE_2_78	0	0	0	6
10761	454_RIF_1484	A	PF761_RIF_AC_16_MR_20	5	0	0	0
10761	454_RIF_1491	A	PF761_RIF_AC_00_GE_2_82	0	0	0	27
10761	454_RIF_1507	A	PF761_RIF_AC_00_GE_2_80	0	0	0	8
10761	454_RIF_1545	A	PF761_RIF_AC_00_GE_2_90	0	0	0	14
10761	454_RIF_1578	A	PF761_RIF_AC_00_GE_2_96	0	0	0	13
10761	454_RIF_1592	A	PF761_RIF_AC_00_GE_2_97	0	0	0	2
10761	454_RIF_1618	A	PF761_RIF_AC_12_MR_11	4	0	0	0
10761	454_RIF_1657	A	PF761_RIF_AC_00_GE_2_105	0	0	0	3
10761	454_RIF_1680	A	PF761_RIF_AC_00_GE_2_107	0	0	0	4
10761	454_RIF_1682	A	PF761_RIF_AC_12_MR_16	10	0	0	0
10761	454_RIF_1734	A	PF761_RIF_AC_00_GE_2_112	0	0	0	3
10761	454_RIF_1757	A	PF761_RIF_AC_00_GE_2_114	0	0	0	9
10761	454_RIF_1772	A	PF761_RIF_AC_00_GE_2_116	0	0	0	5
10761	454_RIF_1821	A	PF761_RIF_AC_34_SZ_2_119	0	0	8	0
10761	454_RIF_1890	A	PF761_RIF_AC_34_SZ_2_123	0	0	18	0
10761	454_RIF_0820	AX	PF761_RIF_AC_16_MR_21	11	0	42	18
10761	454_RIF_1015	AX	PF761_RIF_AC_34_SZ_2_81	0	0	9	0
10761	454_RIF_1029	AX	PF761_RIF_AC_00_GE_2_75	0	0	0	11
10761	454_RIF_1037	AX	PF761_RIF_AC_34_SZ_2_87	0	0	6	0
10761	454_RIF_1060	AX	PF761_RIF_AC_12_MR_10	4	0	0	0
10761	454_RIF_1521	AX	PF761_RIF_AC_00_GE_2_88	0	0	0	3
10761	454_RIF_1560	AX	PF761_RIF_AC_34_SZ_2_100	0	0	3	0
10761	454_RIF_1569	AX	PF761_RIF_AC_16_MR_25	9	0	0	0
10761	454_RIF_1609	AX	PF761_RIF_AC_34_SZ_2_106	0	0	6	0
10761	454_RIF_1629	AX	PF761_RIF_AC_00_GE_2_102	0	0	0	3
10761	454_RIF_1651	AX	PF761_RIF_AC_00_GE_2_104	0	0	0	9
10761	454_RIF_1674	AX	PF761_RIF_AC_00_GE_2_106	0	0	0	7
10761	454_RIF_1761	AX	PF761_RIF_AC_16_MR_27	7	0	0	0
10761	454_RIF_1834	AX	PF761_RIF_AC_60_ER_7	0	0	5	0
10761	454_RIF_1850	AX	PF761_RIF_AC_00_GE_2_125	0	0	0	3
10761	454_RIF_1895	AX	PF761_RIF_AC_00_GE_2_132	0	0	0	8

Isolate	Cluster	A/B	Gene ID	R	LT	S-ER2	G
10761	454_RIF_1926	AX	PF761_RIF_AC_12_MR_24_	7	0	0	0
10761	454_RIF_1935	AX	PF761_RIF_AC_00_GE_2_134	0	0	0	5
10761	454_RIF_0006	B	PF761_RIF_AC_34_SZ_2_82	0	0	11	0
10761	454_RIF_0007	B	PF761_RIF_AC_34_SZ_2_68	0	0	4	0
10761	454_RIF_0034	B	PF761_RIF_AC_34_SZ_2_56	0	0	15	0
10761	454_RIF_0043	B	PF761_RIF_AC_16_MR_11	17	9	25	5
10761	454_RIF_0054	B	PF761_RIF_AC_34_SZ_2_120	0	0	19	0
10761	454_RIF_0060	B	PF761_RIF_AC_34_SZ_2_59	0	0	71	48
10761	454_RIF_0069	B	PF761_RIF_AC_34_SZ_2_33	0	0	10	22
10761	454_RIF_0263	B	PF761_RIF_AC_34_SZ_2_16	0	0	7	0
10761	454_RIF_0277	B	PF761_RIF_AC_34_SZ_2_38	0	0	15	0
10761	454_RIF_0280	B	PF761_RIF_AC_34_SZ_2_32	6	0	42	79
10761	454_RIF_0314	B	PF761_RIF_AC_00_GE_2_30	0	0	6	10
10761	454_RIF_0315	B	PF761_RIF_AC_34_SZ_2_96	0	0	11	0
10761	454_RIF_0342	B	PF761_RIF_AC_00_GE_2_67	0	0	0	7
10761	454_RIF_0419	B	PF761_RIF_AC_00_GE_2_65	0	0	10	9
10761	454_RIF_0449	B	PF761_RIF_AC_12_MR_1	7	0	13	33
10761	454_RIF_0519	B	PF761_RIF_AC_34_SZ_2_85	0	0	6	0
10761	454_RIF_0614	B	PF761_RIF_AC_34_SZ_2_128	0	0	56	0
10761	454_RIF_0673	B	PF761_RIF_AC_34_SZ_2_60	8	4	26	37
10761	454_RIF_0676	B	PF761_RIF_AC_34_SZ_2_61	0	0	7	0
10761	454_RIF_0680	B	PF761_RIF_AC_00_GE_2_81	0	0	37	6
10761	454_RIF_0726	B	PF761_RIF_AC_00_GE_2_14	0	0	4	21
10761	454_RIF_0727	B	PF761_RIF_AC_00_GE_2_15	0	0	0	15
10761	454_RIF_0753	B	PF761_RIF_AC_34_SZ_2_29	0	0	5	0
10761	454_RIF_0761	B	PF761_RIF_AC_34_SZ_2_89	0	0	6	0
10761	454_RIF_0793	B	PF761_RIF_AC_34_SZ_2_75	0	0	7	0
10761	454_RIF_0800	B	PF761_RIF_AC_34_SZ_2_114	0	0	15	0
10761	454_RIF_0804	B	PF761_RIF_AC_00_GE_2_87	0	0	4	12
10761	454_RIF_0805	B	PF761_RIF_AC_00_GE_2_56	0	0	16	12
10761	454_RIF_0807	B	PF761_RIF_AC_00_GE_2_53	0	0	8	29
10761	454_RIF_0825	B	PF761_RIF_AC_34_SZ_2_124	0	0	12	0
10761	454_RIF_0890	B	PF761_RIF_AC_00_GE_2_33	0	0	0	16
10761	454_RIF_0918	B	PF761_RIF_AC_00_GE_2_25	0	0	0	11
10761	454_RIF_0930	B	PF761_RIF_AC_00_GE_2_27	0	0	6	6
10761	454_RIF_0941	B	PF761_RIF_AC_34_SZ_2_116	0	0	3	0
10761	454_RIF_0970	B	PF761_RIF_AC_34_SZ_2_76	0	0	16	0
10761	454_RIF_0976	B	PF761_RIF_AC_00_GE_2_45	0	0	12	23
10761	454_RIF_0986	B	PF761_RIF_AC_00_GE_2_50	0	0	0	26
10761	454_RIF_0989	B	PF761_RIF_AC_12_MR_6	8	0	0	0
10761	454_RIF_0994	B	PF761_RIF_AC_34_SZ_2_57	0	0	14	19
10761	454_RIF_0999	B	PF761_RIF_AC_00_GE_2_58	0	3	0	14
10761	454_RIF_1007	B	PF761_RIF_AC_00_GE_2_62	0	0	0	8
10761	454_RIF_1018	B	PF761_RIF_AC_00_GE_2_70	0	0	8	19
10761	454_RIF_1019	B	PF761_RIF_AC_34_SZ_2_77	0	0	4	0
10761	454_RIF_1030	B	PF761_RIF_AC_26_LT_6	0	7	7	0
10761	454_RIF_1039	B	PF761_RIF_AC_00_GE_2_74	0	0	0	15
10761	454_RIF_1219	B	PF761_RIF_AC_34_SZ_2_27	0	0	6	0
10761	454_RIF_1233	B	PF761_RIF_AC_00_GE_2_26	0	0	0	21
10761	454_RIF_1303	B	PF761_RIF_AC_00_GE_2_40	0	0	0	25
10761	454_RIF_1305	B	PF761_RIF_AC_34_SZ_2_48	0	0	13	0
10761	454_RIF_1331	B	PF761_RIF_AC_00_GE_2_47	0	0	0	31
10761	454_RIF_1353	B	PF761_RIF_AC_34_SZ_2_58	0	0	8	0
10761	454_RIF_1386	B	PF761_RIF_AC_00_GE_2_51	0	0	0	5
10761	454_RIF_1410	B	PF761_RIF_AC_00_GE_2_64	0	0	0	7
10761	454_RIF_1417	B	PF761_RIF_AC_34_SZ_2_66	0	0	5	0
10761	454_RIF_1464	B	PF761_RIF_AC_00_GE_2_76	0	0	0	20
10761	454_RIF_1477	B	PF761_RIF_AC_34_SZ_2_83	4	0	15	7
10761	454_RIF_1522	B	PF761_RIF_AC_34_SZ_2_95	0	0	40	0
10761	454_RIF_1720	B	PF761_RIF_AC_00_GE_2_123_	0	0	0	2
10761	454_RIF_1732	B	PF761_RIF_AC_00_GE_2_111	0	0	0	4
10761	454_RIF_1800	B	PF761_RIF_AC_00_GE_2_115	0	0	0	3
10761	454_RIF_1816	B	PF761_RIF_AC_34_SZ_2_117	0	0	10	0
10761	454_RIF_1889	B	PF761_RIF_AC_00_GE_2_130	0	0	0	5
10761	454_RIF_1900	B	PF761_RIF_AC_34_SZ_2_125	0	0	8	0
10761	454_RIF_1541	BX	PF761_RIF_AC_34_SZ_2_98	0	0	14	0
10761	454_RIF_1949	BX	PF761_RIF_AC_00_GE_2_133	0	0	0	8

Isolate	Cluster	A/B	Gene ID	ER	RN!	MR!	LR!	LRET	ET!	G
10814	454_RIF_0001	A	PF814_RIF_AC_00_GE_2_43	0	0	0	0	10	0	42
10814	454_RIF_0002	A	PF814_RIF_AC_00_GE_2_32	191	9	7	0	10	5	24
10814	454_RIF_0005	A	PF814_RIF_AC_00_GE_2_38	0	0	0	0	0	0	26
10814	454_RIF_0013	A	PF814_RIF_AC_00_GE_2_36	0	0	0	0	0	0	23
10814	454_RIF_0033	A	PF814_RIF_AC_10_ER_42	6	0	0	0	0	0	0
10814	454_RIF_0038	A	PF814_RIF_AC_00_GE_2_141	0	0	0	0	0	0	16
10814	454_RIF_0041	A	PF814_RIF_AC_26_LR_9	0	0	0	4	0	0	0
10814	454_RIF_0071	A	PF814_RIF_AC_12_RN_20	0	4	0	0	0	0	0
10814	454_RIF_0090	A	PF814_RIF_AC_00_GE_2_33	0	8	6	0	0	0	35
10814	454_RIF_0123	A	PF814_RIF_AC_00_GE_2_56	0	0	0	0	0	0	14
10814	454_RIF_0142	A	PF814_RIF_AC_00_GE_2_19	0	0	0	0	4	0	21
10814	454_RIF_0148	A	PF814_RIF_AC_20_PR_2_49	0	0	0	0	4	0	0
10814	454_RIF_0193	A	PF814_RIF_AC_00_GE_2_174	0	0	0	0	0	0	7
10814	454_RIF_0197	A	PF814_RIF_AC_00_GE_2_108	0	0	0	0	0	0	6
10814	454_RIF_0218	A	PF814_RIF_AC_00_GE_2_162	0	0	0	0	0	0	7
10814	454_RIF_0232	A	PF814_RIF_AC_12_RN_23	0	5	0	0	0	0	0
10814	454_RIF_0261	A	PF814_RIF_AC_00_GE_2_114	0	0	0	0	0	0	10
10814	454_RIF_0266	A	PF814_RIF_AC_00_GE_2_110	0	0	0	0	0	0	32
10814	454_RIF_0272	A	PF814_RIF_AC_00_GE_2_142	0	0	0	0	0	0	6
10814	454_RIF_0280	A	PF814_RIF_AC_10_ER_45	2	0	0	0	0	0	0
10814	454_RIF_0352	A	PF814_RIF_AC_10_ER_1	340	12	22	0	8	13	20
10814	454_RIF_0355	A	PF814_RIF_AC_12_RN_22	0	14	0	0	0	0	0
10814	454_RIF_0387	A	PF814_RIF_AC_16_MR_5	29	17	32	18	596	15	56
10814	454_RIF_0397	A	PF814_RIF_AC_00_GE_2_57	25	24	22	10	379	12	21
10814	454_RIF_0401	A	PF814_RIF_AC_00_GE_2_66	0	0	0	0	0	0	21
10814	454_RIF_0447	A	PF814_RIF_AC_16_MR_3	10	0	25	0	346	25	13
10814	454_RIF_0461	A	PF814_RIF_AC_46_ET_37	0	0	0	0	0	1	0
10814	454_RIF_0502	A	PF814_RIF_AC_46_ET_43	0	0	0	0	0	2	0
10814	454_RIF_0536	A	PF814_RIF_AC_00_GE_2_149	0	0	0	0	0	0	5
10814	454_RIF_0548	A	PF814_RIF_AC_10_ER_17	29	21	0	0	7	0	40
10814	454_RIF_0580	A	PF814_RIF_AC_10_ER_2	479	0	13	0	8	9	30
10814	454_RIF_0598	A	PF814_RIF_AC_20_PR_2_34	0	0	0	0	2	0	0
10814	454_RIF_0603	A	PF814_RIF_AC_00_GE_2_116	0	0	0	0	0	0	14
10814	454_RIF_0624	A	PF814_RIF_AC_10_ER_14	31	2	0	0	3	0	26
10814	454_RIF_0629	A	PF814_RIF_AC_00_GE_2_37	0	0	0	0	0	0	11
10814	454_RIF_0642	A	PF814_RIF_AC_00_GE_2_53	13	7	0	0	13	0	31
10814	454_RIF_0646	A	PF814_RIF_AC_20_PR_2_26	0	0	0	0	4	0	0
10814	454_RIF_0656	A	PF814_RIF_AC_46_ET_7	119	0	11	0	0	17	12
10814	454_RIF_0663	A	PF814_RIF_AC_20_PR_2_16	9	0	0	0	8	0	44
10814	454_RIF_0710	A	PF814_RIF_AC_10_ER_9	11	6	0	0	0	0	5
10814	454_RIF_0711	A	PF814_RIF_AC_10_ER_8	99	0	0	0	0	3	7
10814	454_RIF_0715	A	PF814_RIF_AC_20_PR_2_30	0	0	0	0	4	0	0
10814	454_RIF_0729	A	PF814_RIF_AC_20_PR_2_3	0	0	0	0	5	0	26
10814	454_RIF_0769	A	PF814_RIF_AC_00_GE_2_61	0	0	7	0	42	0	23
10814	454_RIF_0843	A	PF814_RIF_AC_00_GE_2_11	12	0	0	0	0	0	8
10814	454_RIF_0844	A	PF814_RIF_AC_10_ER_4	88	0	0	0	0	0	8
10814	454_RIF_0847	A	PF814_RIF_AC_10_ER_12	100	0	0	0	0	2	0
10814	454_RIF_0849	A	PF814_RIF_AC_00_GE_2_134	0	0	0	0	0	0	8
10814	454_RIF_0853	A	PF814_RIF_AC_00_GE_2_180	0	0	0	0	0	0	29
10814	454_RIF_0856	A	PF814_RIF_AC_00_GE_2_7	3	0	0	0	0	0	10
10814	454_RIF_0860	A	PF814_RIF_AC_00_GE_2_14	0	0	0	0	2	0	5
10814	454_RIF_0870	A	PF814_RIF_AC_00_GE_2_26	119	0	0	0	0	0	19
10814	454_RIF_0878	A	PF814_RIF_AC_00_GE_2_12	0	0	0	0	4	0	9
10814	454_RIF_0879	A	PF814_RIF_AC_10_ER_10	26	0	0	0	0	0	0
10814	454_RIF_0884	A	PF814_RIF_AC_00_GE_2_13	44	0	0	0	0	0	20
10814	454_RIF_0889	A	PF814_RIF_AC_00_GE_2_17	0	0	0	0	0	0	12
10814	454_RIF_0905	A	PF814_RIF_AC_00_GE_2_35	15	0	0	0	0	0	27
10814	454_RIF_0907	A	PF814_RIF_AC_00_GE_2_27	0	0	0	0	0	5	39
10814	454_RIF_0928	A	PF814_RIF_AC_00_GE_2_44	10	0	0	0	0	0	55

Isolate	Cluster	A/B	Gene ID	ER	RN!	MR!	LR!	LRET	ET!	G
10814	454_RIF_0943	A	PF814_RIF_AC_00_GE_2_42	0	0	0	0	0	0	55
10814	454_RIF_0947	A	PF814_RIF_AC_00_GE_2_50	0	0	0	0	9	0	19
10814	454_RIF_0952	A	PF814_RIF_AC_46_ET_34	0	0	0	0	0	8	0
10814	454_RIF_0958	A	PF814_RIF_AC_00_GE_2_58	0	0	0	0	30	0	48
10814	454_RIF_0965	A	PF814_RIF_AC_00_GE_2_55	0	0	0	0	2	0	18
10814	454_RIF_0968	A	PF814_RIF_AC_00_GE_2_60	0	0	0	0	0	3	17
10814	454_RIF_0984	A	PF814_RIF_AC_00_GE_2_79	0	0	0	0	3	0	28
10814	454_RIF_1016	A	PF814_RIF_AC_00_GE_2_100	33	0	0	0	0	0	26
10814	454_RIF_1021	A	PF814_RIF_AC_00_GE_2_106	0	0	0	0	7	0	25
10814	454_RIF_1026	A	PF814_RIF_AC_00_GE_2_111	0	0	0	0	0	6	31
10814	454_RIF_1032	A	PF814_RIF_AC_00_GE_2_112	0	0	0	0	0	10	21
10814	454_RIF_1043	A	PF814_RIF_AC_00_GE_2_120	0	0	0	0	0	2	13
10814	454_RIF_1057	A	PF814_RIF_AC_10_ER_29	11	0	0	0	0	15	0
10814	454_RIF_1071	A	PF814_RIF_AC_46_ET_19	0	2	0	0	0	2	0
10814	454_RIF_1090	A	PF814_RIF_AC_00_GE_2_1	0	0	0	0	0	0	10
10814	454_RIF_1099	A	PF814_RIF_AC_00_GE_2_8	0	0	0	0	0	0	7
10814	454_RIF_1106	A	PF814_RIF_AC_00_GE_2_10	0	0	0	0	0	0	7
10814	454_RIF_1120	A	PF814_RIF_AC_00_GE_2_4	0	0	0	0	0	0	5
10814	454_RIF_1125	A	PF814_RIF_AC_00_GE_2_3	0	0	0	0	0	0	6
10814	454_RIF_1126	A	PF814_RIF_AC_00_GE_2_9	0	0	0	0	0	0	12
10814	454_RIF_1138	A	PF814_RIF_AC_00_GE_2_5	0	0	0	0	0	0	17
10814	454_RIF_1146	A	PF814_RIF_AC_00_GE_2_16	0	0	0	0	0	0	8
10814	454_RIF_1166	A	PF814_RIF_AC_00_GE_2_18	0	0	0	0	0	0	12
10814	454_RIF_1167	A	PF814_RIF_AC_00_GE_2_20	0	0	0	0	0	0	7
10814	454_RIF_1168	A	PF814_RIF_AC_00_GE_2_21	0	0	0	0	0	0	8
10814	454_RIF_1175	A	PF814_RIF_AC_00_GE_2_25	0	0	0	0	0	0	16
10814	454_RIF_1179	A	PF814_RIF_AC_12_RN_3	0	7	0	0	0	0	0
10814	454_RIF_1180	A	PF814_RIF_AC_00_GE_2_30	0	0	0	0	0	0	7
10814	454_RIF_1191	A	PF814_RIF_AC_20_PR_2_7	0	0	0	0	140	0	0
10814	454_RIF_1193	A	PF814_RIF_AC_00_GE_2_24	0	0	0	0	0	0	6
10814	454_RIF_1202	A	PF814_RIF_AC_00_GE_2_29	39	0	0	0	0	0	14
10814	454_RIF_1235	A	PF814_RIF_AC_00_GE_2_46	0	0	0	0	0	0	21
10814	454_RIF_1238	A	PF814_RIF_AC_00_GE_2_52	0	0	0	0	0	0	15
10814	454_RIF_1252	A	PF814_RIF_AC_00_GE_2_48	0	0	0	0	0	0	48
10814	454_RIF_1256	A	PF814_RIF_AC_00_GE_2_47	0	0	0	0	0	4	6
10814	454_RIF_1267	A	PF814_RIF_AC_00_GE_2_54	0	0	0	0	0	0	5
10814	454_RIF_1279	A	PF814_RIF_AC_00_GE_2_63	0	0	0	0	0	0	9
10814	454_RIF_1289	A	PF814_RIF_AC_00_GE_2_59	0	0	0	0	0	0	6
10814	454_RIF_1301	A	PF814_RIF_AC_00_GE_2_62	0	0	0	0	0	0	5
10814	454_RIF_1314	A	PF814_RIF_AC_00_GE_2_64	0	0	0	0	0	0	17
10814	454_RIF_1333	A	PF814_RIF_AC_00_GE_2_65	0	0	0	0	0	0	31
10814	454_RIF_1337	A	PF814_RIF_AC_10_ER_21	35	0	0	0	0	0	0
10814	454_RIF_1341	A	PF814_RIF_AC_00_GE_2_70	0	0	0	0	0	0	12
10814	454_RIF_1344	A	PF814_RIF_AC_00_GE_2_68	0	0	0	0	0	0	37
10814	454_RIF_1354	A	PF814_RIF_AC_00_GE_2_82	0	0	0	0	0	0	4
10814	454_RIF_1355	A	PF814_RIF_AC_00_GE_2_86	0	0	0	0	0	0	5
10814	454_RIF_1377	A	PF814_RIF_AC_00_GE_2_88	0	0	0	0	0	0	7
10814	454_RIF_1378	A	PF814_RIF_AC_00_GE_2_92	0	0	0	0	0	0	24
10814	454_RIF_1384	A	PF814_RIF_AC_00_GE_2_81	0	0	0	0	0	0	8
10814	454_RIF_1387	A	PF814_RIF_AC_00_GE_2_94	0	0	0	0	0	0	11
10814	454_RIF_1404	A	PF814_RIF_AC_20_PR_2_21	0	0	0	0	8	0	0
10814	454_RIF_1445	A	PF814_RIF_AC_00_GE_2_109	0	0	0	0	0	0	5
10814	454_RIF_1455	A	PF814_RIF_AC_00_GE_2_105	0	0	0	0	0	0	14
10814	454_RIF_1456	A	PF814_RIF_AC_00_GE_2_103	0	0	0	0	0	0	6
10814	454_RIF_1479	A	PF814_RIF_AC_00_GE_2_115	0	0	0	0	0	0	33
10814	454_RIF_1481	A	PF814_RIF_AC_00_GE_2_113	0	0	0	0	0	0	10
10814	454_RIF_1492	A	PF814_RIF_AC_00_GE_2_117	0	0	0	0	0	0	11
10814	454_RIF_1504	A	PF814_RIF_AC_00_GE_2_119	0	0	0	0	0	0	12
10814	454_RIF_1518	A	PF814_RIF_AC_10_ER_28	9	0	0	0	0	0	0

Isolate	Cluster	A/B	Gene ID	ER	RN!	MR!	LR!	LRET	ET!	G
10814	454_RIF_1537	A	PF814_RIF_AC_00_GE_2_127	0	0	0	0	0	0	25
10814	454_RIF_1570	A	PF814_RIF_AC_46_ET_16	0	0	0	0	0	7	0
10814	454_RIF_1600	A	PF814_RIF_AC_00_GE_2_136	0	0	0	0	0	0	20
10814	454_RIF_1605	A	PF814_RIF_AC_10_ER_31	9	0	0	0	0	0	0
10814	454_RIF_1606	A	PF814_RIF_AC_12_RN_10	0	5	0	0	0	0	0
10814	454_RIF_1619	A	PF814_RIF_AC_10_ER_30	12	0	0	0	0	0	0
10814	454_RIF_1620	A	PF814_RIF_AC_00_GE_2_135	0	0	0	0	0	0	12
10814	454_RIF_1633	A	PF814_RIF_AC_12_RN_12	0	3	0	0	0	0	0
10814	454_RIF_1642	A	PF814_RIF_AC_12_RN_11	0	8	0	0	0	0	0
10814	454_RIF_1661	A	PF814_RIF_AC_00_GE_2_143	0	0	0	0	0	0	5
10814	454_RIF_1670	A	PF814_RIF_AC_20_PR_2_39	0	0	0	0	6	0	0
10814	454_RIF_1681	A	PF814_RIF_AC_10_ER_33	15	0	0	0	0	0	0
10814	454_RIF_1683	A	PF814_RIF_AC_10_ER_34	11	0	0	0	0	0	0
10814	454_RIF_1690	A	PF814_RIF_AC_00_GE_2_144	0	0	0	0	0	0	8
10814	454_RIF_1691	A	PF814_RIF_AC_00_GE_2_148	0	0	0	0	0	0	12
10814	454_RIF_1703	A	PF814_RIF_AC_46_ET_25	0	0	0	0	0	2	0
10814	454_RIF_1708	A	PF814_RIF_AC_00_GE_2_151	0	0	0	0	0	0	4
10814	454_RIF_1712	A	PF814_RIF_AC_00_GE_2_153	0	0	0	0	0	0	13
10814	454_RIF_1728	A	PF814_RIF_AC_12_RN_17	0	2	0	0	0	0	0
10814	454_RIF_1751	A	PF814_RIF_AC_10_ER_36	7	0	0	0	0	0	0
10814	454_RIF_1756	A	PF814_RIF_AC_00_GE_2_158	0	0	0	0	0	0	10
10814	454_RIF_1762	A	PF814_RIF_AC_00_GE_2_157	0	0	0	0	0	0	6
10814	454_RIF_1764	A	PF814_RIF_AC_00_GE_2_160	0	0	0	0	0	0	7
10814	454_RIF_1769	A	PF814_RIF_AC_00_GE_2_155	0	0	0	0	0	0	6
10814	454_RIF_1775	A	PF814_RIF_AC_20_PR_2_44	0	0	0	0	7	0	0
10814	454_RIF_1788	A	PF814_RIF_AC_00_GE_2_161	0	0	0	0	0	0	8
10814	454_RIF_1789	A	PF814_RIF_AC_00_GE_2_163	0	0	0	0	0	0	9
10814	454_RIF_1805	A	PF814_RIF_AC_26_LR_10	0	0	0	0	0	0	0
10814	454_RIF_1820	A	PF814_RIF_AC_00_GE_2_164	0	0	0	0	0	0	10
10814	454_RIF_1824	A	PF814_RIF_AC_10_ER_38	8	0	0	0	0	0	0
10814	454_RIF_1825	A	PF814_RIF_AC_00_GE_2_168	0	0	0	0	0	0	3
10814	454_RIF_1842	A	PF814_RIF_AC_00_GE_2_166	0	0	0	0	0	0	6
10814	454_RIF_1854	A	PF814_RIF_AC_10_ER_40	5	0	0	0	0	0	0
10814	454_RIF_1875	A	PF814_RIF_AC_46_ET_38	0	0	0	0	0	11	0
10814	454_RIF_1882	A	PF814_RIF_AC_00_GE_2_176	0	0	0	0	0	0	14
10814	454_RIF_1911	A	PF814_RIF_AC_10_ER_44	6	0	0	0	0	0	0
10814	454_RIF_1049	AX	PF814_RIF_AC_00_GE_2_123	0	0	0	0	0	0	11
10814	454_RIF_1458	AX	PF814_RIF_AC_00_GE_2_102	0	0	0	0	0	0	9
10814	454_RIF_1485	AX	PF814_RIF_AC_10_ER_25	11	0	0	0	0	0	0
10814	454_RIF_1498	AX	PF814_RIF_AC_46_ET_12	0	0	0	0	0	9	0
10814	454_RIF_1523	AX	PF814_RIF_AC_00_GE_2_126	0	0	0	0	0	0	18
10814	454_RIF_1539	AX	PF814_RIF_AC_00_GE_2_128	0	0	0	0	0	0	6
10814	454_RIF_1548	AX	PF814_RIF_AC_20_PR_2_28	0	0	0	0	2	0	0
10814	454_RIF_1549	AX	PF814_RIF_AC_00_GE_2_125	0	0	0	0	0	0	10
10814	454_RIF_1591	AX	PF814_RIF_AC_00_GE_2_133	0	0	0	0	0	0	8
10814	454_RIF_1603	AX	PF814_RIF_AC_00_GE_2_137	0	0	0	0	0	0	11
10814	454_RIF_1621	AX	PF814_RIF_AC_10_ER_32	6	0	0	0	0	0	0
10814	454_RIF_1622	AX	PF814_RIF_AC_20_PR_2_35	0	0	0	0	6	0	0
10814	454_RIF_1675	AX	PF814_RIF_AC_00_GE_2_145	0	0	0	0	0	0	6
10814	454_RIF_1677	AX	PF814_RIF_AC_00_GE_2_146	0	0	0	0	0	0	9
10814	454_RIF_1715	AX	PF814_RIF_AC_46_ET_28	0	0	0	0	0	5	0
10814	454_RIF_1773	AX	PF814_RIF_AC_00_GE_2_159	0	0	0	0	0	0	3
10814	454_RIF_1840	AX	PF814_RIF_AC_00_GE_2_171	0	0	0	0	0	0	3
10814	454_RIF_1851	AX	PF814_RIF_AC_10_ER_41	7	0	0	0	0	0	0
10814	454_RIF_1906	AX	PF814_RIF_AC_00_GE_2_178	0	0	0	0	0	0	10
10814	454_RIF_1938	AX	PF814_RIF_AC_00_GE_2_181	0	0	0	0	0	0	2
10814	454_RIF_0006	B	PF814_RIF_AC_00_GE_2_69	0	0	0	0	0	0	24
10814	454_RIF_0025	B	PF814_RIF_AC_00_GE_2_140	0	0	0	0	0	15	24
10814	454_RIF_0060	B	PF814_RIF_AC_00_GE_2_98	0	0	0	0	0	0	13

Isolate	Cluster	A/B	Gene ID	ER	RN!	MR!	LR!	LRET	ET!	G
10814	454_RIF_0094	B	PF814_RIF_AC_20_PR_2_48	0	0	0	0	4	0	128
10814	454_RIF_0101	B	PF814_RIF_AC_00_GE_2_90	0	0	0	0	0	0	24
10814	454_RIF_0116	B	PF814_RIF_AC_00_GE_2_39	0	0	0	0	0	0	6
10814	454_RIF_0203	B	PF814_RIF_AC_00_GE_2_80	0	0	0	0	0	0	55
10814	454_RIF_0237	B	PF814_RIF_AC_00_GE_2_132	0	0	0	0	0	0	9
10814	454_RIF_0239	B	PF814_RIF_AC_00_GE_2_83	0	0	0	0	0	0	9
10814	454_RIF_0240	B	PF814_RIF_AC_00_GE_2_124	0	0	0	0	0	0	7
10814	454_RIF_0241	B	PF814_RIF_AC_46_ET_10	0	17	24	0	8	10	62
10814	454_RIF_0287	B	PF814_RIF_AC_00_GE_2_172	0	0	0	0	0	0	10
10814	454_RIF_0289	B	PF814_RIF_AC_20_PR_2_25	0	0	0	0	3	0	0
10814	454_RIF_0314	B	PF814_RIF_AC_00_GE_2_121	0	0	0	0	0	0	16
10814	454_RIF_0317	B	PF814_RIF_AC_10_ER_22	9	0	0	0	0	0	0
10814	454_RIF_0380	B	PF814_RIF_AC_00_GE_2_173	0	0	0	0	0	0	44
10814	454_RIF_0561	B	PF814_RIF_AC_00_GE_2_77	0	0	2	0	0	0	56
10814	454_RIF_0562	B	PF814_RIF_AC_16_MR_8	0	0	18	20	209	3	37
10814	454_RIF_0648	B	PF814_RIF_AC_20_PR_2_27	0	0	0	0	2	0	0
10814	454_RIF_0671	B	PF814_RIF_AC_00_GE_2_93	0	0	6	0	8	0	32
10814	454_RIF_0673	B	PF814_RIF_AC_00_GE_2_78	0	0	0	0	8	0	29
10814	454_RIF_0677	B	PF814_RIF_AC_00_GE_2_85	0	8	6	0	0	5	22
10814	454_RIF_0727	B	PF814_RIF_AC_20_PR_2_4	0	0	0	0	7	0	32
10814	454_RIF_0728	B	PF814_RIF_AC_00_GE_2_51	0	0	0	0	0	91	37
10814	454_RIF_0742	B	PF814_RIF_AC_00_GE_2_118	0	0	0	0	0	0	5
10814	454_RIF_0751	B	PF814_RIF_AC_00_GE_2_175	0	0	0	0	0	0	13
10814	454_RIF_0796	B	PF814_RIF_AC_00_GE_2_73	0	0	4	0	12	0	31
10814	454_RIF_0809	B	PF814_RIF_AC_20_PR_2_24	0	0	0	0	8	0	0
10814	454_RIF_0825	B	PF814_RIF_AC_46_ET_18	0	0	0	0	0	34	44
10814	454_RIF_0891	B	PF814_RIF_AC_46_ET_3	0	0	0	0	0	24	0
10814	454_RIF_0895	B	PF814_RIF_AC_46_ET_31	0	0	0	0	0	51	0
10814	454_RIF_0897	B	PF814_RIF_AC_00_GE_2_23	0	0	0	0	19	0	8
10814	454_RIF_0898	B	PF814_RIF_AC_20_PR_2_2	0	0	0	0	6	0	23
10814	454_RIF_0950	B	PF814_RIF_AC_00_GE_2_49	0	0	3	0	0	0	8
10814	454_RIF_0957	B	PF814_RIF_AC_12_RN_13	0	4	0	0	0	0	0
10814	454_RIF_0996	B	PF814_RIF_AC_00_GE_2_97	0	0	0	0	0	0	33
10814	454_RIF_1008	B	PF814_RIF_AC_00_GE_2_95	0	0	0	0	0	0	10
10814	454_RIF_1020	B	PF814_RIF_AC_00_GE_2_104	0	0	0	0	0	3	21
10814	454_RIF_1067	B	PF814_RIF_AC_00_GE_2_138	0	0	0	0	0	9	12
10814	454_RIF_1081	B	PF814_RIF_AC_46_ET_35	0	0	0	0	0	5	16
10814	454_RIF_1194	B	PF814_RIF_AC_00_GE_2_28	0	0	0	0	0	0	12
10814	454_RIF_1199	B	PF814_RIF_AC_00_GE_2_31	0	0	0	0	0	0	15
10814	454_RIF_1242	B	PF814_RIF_AC_00_GE_2_41	0	0	0	0	0	0	23
10814	454_RIF_1346	B	PF814_RIF_AC_00_GE_2_72	0	0	0	0	0	0	30
10814	454_RIF_1360	B	PF814_RIF_AC_00_GE_2_75	0	0	0	0	0	0	22
10814	454_RIF_1361	B	PF814_RIF_AC_00_GE_2_87	0	0	0	0	0	0	9
10814	454_RIF_1373	B	PF814_RIF_AC_00_GE_2_71	0	0	0	0	0	0	13
10814	454_RIF_1379	B	PF814_RIF_AC_00_GE_2_89	0	0	0	0	0	0	16
10814	454_RIF_1381	B	PF814_RIF_AC_00_GE_2_74	0	0	0	0	0	0	5
10814	454_RIF_1390	B	PF814_RIF_AC_00_GE_2_84	0	0	0	0	0	0	11
10814	454_RIF_1433	B	PF814_RIF_AC_00_GE_2_99	0	0	0	0	0	0	8
10814	454_RIF_1477	B	PF814_RIF_AC_00_GE_2_130	0	0	0	0	0	16	14
10814	454_RIF_1538	B	PF814_RIF_AC_46_ET_15	0	0	0	0	0	17	0
10814	454_RIF_1565	B	PF814_RIF_AC_20_PR_2_32	0	0	0	0	4	0	0
10814	454_RIF_1599	B	PF814_RIF_AC_16_MR_15	0	0	3	0	0	0	0
10814	454_RIF_1720	B	PF814_RIF_AC_00_GE_2_154	0	0	0	0	0	0	4
10814	454_RIF_1929	B	PF814_RIF_AC_46_ET_44	0	0	0	0	0	18	0
10814	454_RIF_1530	BX	PF814_RIF_AC_46_ET_14	0	0	0	0	0	6	0
10814	454_RIF_1588	BX	PF814_RIF_AC_00_GE_2_131	0	0	0	0	0	13	42
10814	454_RIF_1625	BX	PF814_RIF_AC_00_GE_2_139	0	0	0	0	2	0	77
10814	454_RIF_1628	BX	PF814_RIF_AC_20_PR_2_37	0	0	0	0	4	0	0
10814	454_RIF_1947	X	PF814_RIF_AC_00_GE_2_179	0	0	0	0	0	0	8

Appendix Table 9.12. *Raw stevor read counts from 454 amplicon sequencing of cDNA and gDNA from patient iolates that were put under differential selection pressure. AC (acute), CU (culture-adapted) HB (HBEC-slelected) R+ (rosetting clone), R- (non-rosetting fraction). The data in the table was generated by mapping 454 reads onto homologous genomes and computing read counts using Artemis. An average read count was taken when samples were sequenced in duplicates or triplicates. Genes in the table had greater than 5X read coverage.*

Isolate	Gene ID	AC_R	AC_T	AC_G	CU_R	CU_LRET	CU_T	CU_S	CU_G
9215	PF8443_8_4_U533600	431	138	62	264	510	3324	5283	61
9215	PF8443_8_4_U026200	1063	1870	116	2116	905	801	444	338
9215	PF8443_8_4_U516100	2147	213	825	265	371	492	285	1065
9215	PF8443_8_4_U265000	2278	288	582	662	342	341	178	969
9215	PF8443_8_4_U546300	2152	173	374	725	772	426	251	585
9215	PF8443_8_4_U511100	709	298	916	390	555	631	448	1477
9215	PF8443_8_4_U523400	1126	250	715	609	320	225	153	1198
9215	PF8443_8_4_U503400	1937	597	225	657	347	283	126	416
9215	PF8443_8_4_U559100	1552	160	194	545	174	167	497	350
9215	PF8443_8_4_U537000	882	294	382	260	194	387	325	580
9215	PF8443_8_4_U560500	724	117	958	110	79	148	98	900
9215	PF8443_8_4_U550600	786	79	379	212	227	265	139	821
9215	PF8443_8_4_U505800	876	175	296	425	285	271	171	369
9215	PF8443_8_4_U542800	389	52	457	55	63	159	446	874
9215	PF8443_8_4_U503300	625	108	283	385	111	256	271	266
9215	PF8443_8_4_U559700	995	50	148	197	102	97	64	249
9215	PF8443_8_4_U556500	709	81	254	65	45	82	116	235
9215	PF8443_8_4_U530700	166	50	116	191	54	66	57	160
9215	PF8443_8_4_U540400	568	36	72	3	5	1	7	1
9215	PF8443_8_4_U518800	133	18	61	30	15	60	66	103
9215	PF8443_8_4_U546700	56	55	12	21	19	12	7	35

Isolate	Gene ID	AC_R	AC_G	CU_R	CU_LRET	CU_T	CU_S	CU_G	HB_R	HB_LRET	HB_T	HB_S	HB_G
9106	PF8443_8_7_U557100	343	987	572	121	249	132	1894	702	3319	3733	3033	1867
9106	PF8443_8_7_U563400	353	117	3807	3825	5099	3785	123	471	1134	1429	2240	169
9106	PF8443_8_7_U503200	525	1224	1296	1223	777	523	2220	133	343	198	191	1567
9106	PF8443_8_7_U449600	146	616	266	314	409	257	1024	34	63	285	406	964
9106	PF8443_8_7_U447100	73	410	276	187	546	485	791	9	39	532	184	844
9106	PF8443_8_7_U449500	103	447	184	270	312	212	672	30	45	247	353	673
9106	PF8443_8_7_U486000	64	242	1191	924	246	32	451	22	43	26	75	435
9106	PF8443_8_7_U543200	80	291	201	215	202	147	528	19	64	110	111	447
9106	PF8443_8_7_U561200	66	343	122	94	67	71	583	21	20	34	46	571
9106	PF8443_8_7_U112400	21	187	60	8	6	4	598	21	22	11	99	621
9106	PF8443_8_7_U541900	120	366	169	25	37	22	564	2	9	5	11	306
9106	PF8443_8_7_U554300	106	202	264	33	10	12	442	43	66	14	75	406
9106	PF8443_8_7_U472600	95	275	83	61	50	27	325	80	128	65	83	366
9106	PF8443_8_7_U436400	24	206	41	60	105	77	449	16	27	32	78	455
9106	PF8443_8_7_U520800	49	252	85	56	89	49	338	9	35	30	51	337
9106	PF8443_8_7_U561400	116	300	222	89	73	39	380	6	16	14	10	37
9106	PF8443_8_7_U468200	27	109	130	27	37	16	107	20	57	10	188	99
9106	PF8443_8_7_U547700	38	132	68	39	67	42	87	16	32	32	41	146
9106	PF8443_8_7_U560500	72	47	0	0	0	0	19	388	31	0	0	48
9106	PF8443_8_7_U557900	20	53	137	20	11	10	137	49	56	2	51	89
9106	PF8443_8_7_U527100	16	43	64	44	15	8	65	19	42	3	54	73
9106	PF8443_8_7_U538400	43	73	27	7	16	7	67	5	4	2	2	84
9106	PF8443_8_7_U555100	8	28	5	8	3	4	42	6	3	4	2	46
9106	PF8443_8_7_U542600	7	16	3	0	2	4	10	0	1	0	2	11
9106	PF8443_8_7_U542500	6	4	7	0	3	2	10	0	0	0	0	4

Isolate	Gene ID	AC_R	AC_T	AC_G	CU_R	CU_LRET	CU_T	CU_S	CU_G	HB_R	HB_LRET	HB_T	HB_S	HB_G
10668	PF8443_8_8_U497600	6	50	74.5	4	17	2	128	40	18	1	4	8	13
10668	PF8443_8_8_U549200	7	0	30	9	23	44	409	54	38	21	18	30	41
10668	PF8443_8_8_U492900	68	3	113	28	51	64	279	133	240	38	53	19	88
10668	PF8443_8_8_U537300	49	27	104	37	76	115	1001	57	12	8	14	98	73
10668	PF8443_8_8_U543100	305	3	147.5	78	101	277	773	87	119	15	32	82	78
10668	PF8443_8_8_U550500	69	10	235.5	50	69	97	729	218	376	25	41	80	202
10668	PF8443_8_8_U538400	169	44	240.5	57	100	142	577	118	302	110	135	66	180
10668	PF8443_8_8_U560900	319	25	144.5	71	222	326	728	120	287	16	26	82	126
10668	PF8443_8_8_U097100	225	17	171.5	97	64	221	908	149	550	49	50	91	127
10668	PF8443_8_8_U540800	258	7	98	62	135	141	1581	95	235	21	23	172	73
10668	PF8443_8_8_U514700	195	8	244	60	78	159	1806	261	422	87	95	206	185
10668	PF8443_8_8_U498500	109	14	158	182	85	99	1435	128	723	293	308	157	120
10668	PF8443_8_8_U537100	255	4	383.5	115	178	360	845	343	602	237	253	79	308
10668	PF8443_8_8_U553100	63	19	220	33	53	152	1199	194	416	663	799	110	159
10668	PF8443_8_8_U561500	153	4	90.5	43	166	1567	1835	118	110	78	144	207	101
10668	PF8443_8_8_U566600	126	22	619	131	166	429	2570	507	1011	296	471	290	612
10668	PF8443_8_8_U563300	1158	57	771	20	14	1	2976	0	385	646	620	348	670
10668	PF8443_8_8_U501200	1014	9	481	2260	604	412	3123	442	2084	282	373	319	600
10668	PF8443_8_8_U553300	2802	13	543	545	1242	1275	5859	505	1045	1219	1225	621	474
10668	PF8443_8_8_U542300	204	1	546	142	200	251	3822	391	9665	1334	1267	418	433
10668	PF8443_8_8_U488800	968	34	744	497	687	882	8400	659	1419	2372	3609	896	711
10668	PF8443_8_8_U535700	791	43	441.5	486	587	1247	14102	446	2063	722	1979	1460	548

Isolate	Gene ID	R+R	R+LRET	R+T	R+S	R+G	R-R	R-LRET	R-T	R-S	R-G
SA075	PFSAO75_493100	47	10	12	25	95	742	4683	5061	7484	118
SA075	PFSAO75_128400	372	2012	2337	510	393	282	191	197	527	408
SA075	PFSAO75_051700	55	23	19	418	85	138	1678	642	1413	100
SA075	PFSAO75_154400	726	35	34	633	308	202	24	83	147	269
SA075	PFSAO75_001200	237	13	32	654	318	245	60	29	136	438
SA075	PFSAO75_258600	279	21	19	146	438	247	164	43	195	503
SA075	PFSAO75_347800	264	24	34	88	389	213	174	79	268	441
SA075	PFSAO75_072200	258	201	89	87	387	156	35	16	85	430
SA075	PFSAO75_348400	215	13	20	92	358	65	72	43	151	496
SA075	PFSAO75_073100	202	17	21	88	353	83	34	17	42	338
SA075	PFSAO75_PN67	173	10	18	27	390	45	37	18	32	385
SA075	PFSAO75_492500	94	80	51	54	172	69	52	42	139	213
SA075	PFSAO75_PN64	88	9	36	144	95	66	40	40	254	115
SA075	PFSAO75_493000	168	4	20	39	269	14	17	7	44	296
SA075	PFSAO75_011100	49	195	101	138	54	25	60	33	41	75
SA075	PFSAO75_071200	84	38	18	38	125	16	17	20	58	166
SA075	PFSAO75_PN66	52	18	3	90	115	55	23	11	35	160
SA075	PFSAO75_PN61	49	2	1	9	126	3	16	8	29	198
SA075	PFSAO75_051200	35	1	2	38	107	27	14	18	33	107
SA075	PFSAO75_PN62	12	33	19	10	42	28	27	52	63	64
SA075	PFSAO75_PN65	27	3	4	78	59	25	2	6	27	74
SA075	PFSAO75_PN60	36	0	4	4	66	37	16	11	31	86
SA075	PFSAO75_PN63	39	23	15	5	7	2	3	2	14	128
SA075	PFSAO75_011900	6	0	1	12	79	49	30	4	40	14
SA075	PFSAO75_492100	2	0	0	0	26	0	3	0	6	34

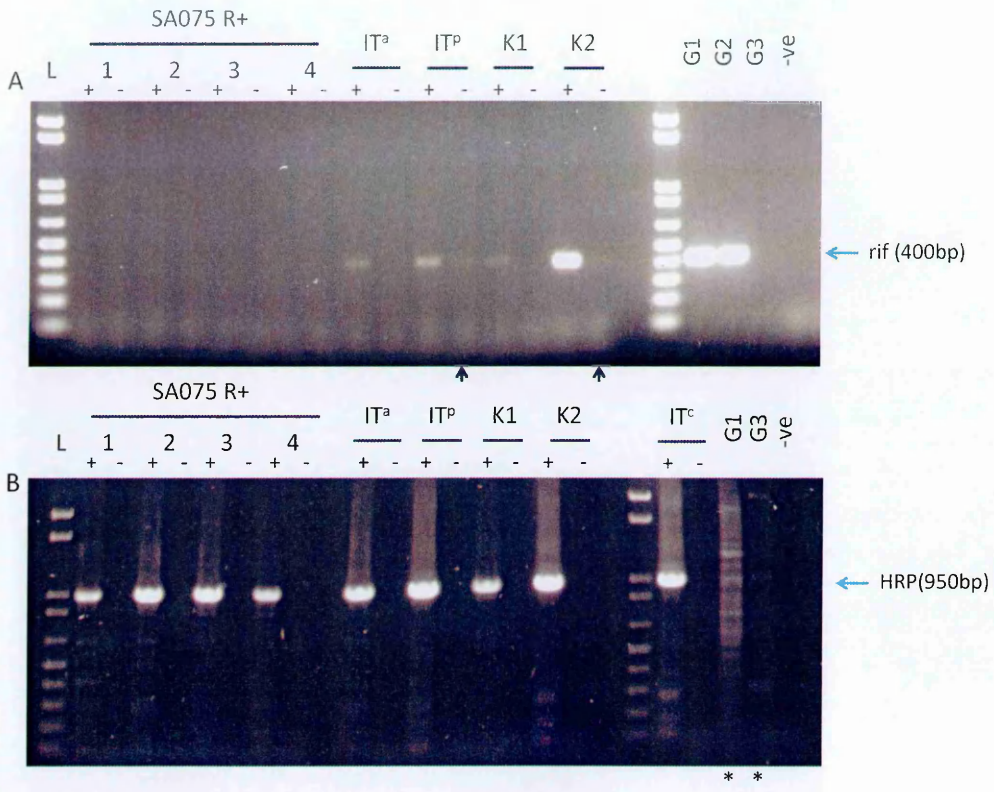
Appendix Table 9.13. *Raw stevor read counts from 454 amplicon sequencing of cDNA and gDNA from a patient isolate, 8383 put under differential selection pressure. AC (acute rings and ex vivo trophozoites) and CU (culture-adapted) and fresh-from-the-arm patient samples that were matured in culture for a single IDC, (acute rings, and ex vivo late rings to schizonts). These isolates lacked whole genome sequence data. Therefore the data in the table was generated by de novo assembly of 454 reads. The numbers of reads that went into contig assembly were computed per sample and these numbers were used at the read count data. !: samples that had limited sequencing depths (less than 100 reads).*

Isolate	Cluster ID	Sequence ID	AC_R	AC_T	AC_G	CU_P	CU_T	CU_S	CU_G
8383	STE_454_001	PF383_STE_CU_06_P1_4	563	9287	233	8	2314	95	47
8383	STE_454_002	PF383_STE_AC_00_R1_3_1	94	175	190	16	24	778	24
8383	STE_454_005	PF383_STE_AC_00_G2_3_8	58	170	206	159	0	169	0
8383	STE_454_012	PF383_STE_CU_30_SZ_4	0	0	0	0	0	6	0
8383	STE_454_019	PF383_STE_AC_00_R1_3_18	43	240	1284	5	49	8	0
8383	STE_454_033	PF383_STE_CU_00_GE_3_2	609	886	879	460	504	20	204
8383	STE_454_050	PF383_STE_CU_18_TR_1	725	1117	754	1328	287	179	157
8383	STE_454_051	PF383_STE_CU_30_SZ_7	198	767	768	609	1667	75	149
8383	STE_454_052	PF383_STE_CU_06_P1_2	26	483	185	57	0	15	22
8383	STE_454_057	PF383_STE_AC_00_R1_3_9	81	1077	382	1257	137	261	99
8383	STE_454_078	PF383_STE_AC_00_G2_3_6	513	176	286	192	7	53	43
8383	STE_454_089	PF383_STE_AC_24_T1_10	89	706	539	763	18	215	119
8383	STE_454_095	PF383_STE_CU_00_GE_3_14	217	666	605	130	14	30	83
8383	STE_454_103	PF383_STE_CU_00_GE_3_9	832	387	1204	0	0	77	343
8383	STE_454_108	PF383_STE_AC_00_G1_3_10	39	469	562	300	33	320	136
8383	STE_454_115	PF383_STE_AC_00_R1_3_10	9	131	103	184	0	848	29
8383	STE_454_148	PF383_STE_CU_06_P1_12	170	881	517	111	75	8	126
8383	STE_454_153	PF383_STE_CU_30_SZ_10	38	113	155	319	12	757	52
8383	STE_454_180	PF383_STE_AC_00_G1_3_4	0	26	28	0	0	0	13
8383	STE_454_200	PF383_STE_AC_00_G2_3_18	0	94	353	0	0	0	40
8383	STE_454_300	PF383_STE_AC_24_T1_22	0	6	0	0	0	0	0

Isolate	Cluster ID	Gene ID	ER!	MR	T	S	G
10727	STE_454_001	PF727_STE_AC_32_S1_3_9	35	300	3904	1177	107
10727	STE_454_002	PF727_STE_AC_12_R1_3_2	0	68	8	45	15
10727	STE_454_005	PF727_STE_AC_00_ER_3_6	5	0	0	0	0
10727	STE_454_008	PF727_STE_AC_12_R1_3_18	0	158	97	109	194
10727	STE_454_026	PF727_STE_AC_00_GE_3_7	0	198	19	75	83
10727	STE_454_030	PF727_STE_AC_00_ER_3_3	28	0	0	0	0
10727	STE_454_033	PF727_STE_AC_00_ER_3_2	17	0	0	0	0
10727	STE_454_034	PF727_STE_AC_12_R1_3_11	0	382	22	108	254
10727	STE_454_037	PF727_STE_AC_00_GE_3_29	0	131	37	89	79
10727	STE_454_044	PF727_STE_AC_00_ER_3_9	30	0	0	0	0
10727	STE_454_047	PF727_STE_AC_00_ER_3_11	43	0	0	0	0
10727	STE_454_050	PF727_STE_AC_00_ER_3_10	44	0	0	0	0
10727	STE_454_051	PF727_STE_AC_00_ER_3_4	34	0	0	0	0
10727	STE_454_052	PF727_STE_AC_12_R1_3_4	0	60	166	83	63
10727	STE_454_061	PF727_STE_AC_12_R2_3_1	0	22	0	0	0
10727	STE_454_064	PF727_STE_AC_32_S1_3_4	0	43	26	28	234
10727	STE_454_066	PF727_STE_AC_00_ER_3_7	26	0	0	0	0
10727	STE_454_073	PF727_STE_AC_22_T1_3_23	0	677	74	414	383
10727	STE_454_089	PF727_STE_AC_00_ER_3_12	17	0	0	0	0
10727	STE_454_091	PF727_STE_AC_12_R1_3_7	0	273	82	266	10
10727	STE_454_099	PF727_STE_AC_00_ER_3_14	15	0	0	0	0
10727	STE_454_100	PF727_STE_AC_12_R1_3_3	0	128	0	49	15
10727	STE_454_101	PF727_STE_AC_12_R1_3_1	0	111	56	156	369
10727	STE_454_103	PF727_STE_AC_00_ER_3_5	111	0	0	0	0
10727	STE_454_107	PF727_STE_AC_22_T2_3_25	0	26	25	8	19
10727	STE_454_109	PF727_STE_AC_22_T1_3_16	0	815	20	273	108
10727	STE_454_121	PF727_STE_AC_32_S1_3_28	0	288	164	170	322
10727	STE_454_122	PF727_STE_AC_12_R1_3_12	53	46	26	79	139
10727	STE_454_150	PF727_STE_AC_12_R1_3_6	0	318	22	246	286
10727	STE_454_155	PF727_STE_AC_22_T1_3_14	0	46	28	93	235
10727	STE_454_156	PF727_STE_AC_22_T1_3_8	0	104	1509	737	174
10727	STE_454_161	PF727_STE_AC_22_T1_3_17	0	319	189	139	318
10727	STE_454_164	PF727_STE_AC_32_S1_3_15	0	611	195	280	195
10727	STE_454_166	PF727_STE_AC_22_T1_3_18	0	635	251	356	212
10727	STE_454_169	PF727_STE_AC_22_T1_3_21	0	445	67	190	191
10727	STE_454_171	PF727_STE_AC_00_GE_3_28	0	153	45	79	243
10727	STE_454_173	PF727_STE_AC_00_GE_3_14	0	0	96	30	559
10727	STE_454_176	PF727_STE_AC_32_S1_3_5	0	116	22	15	24
10727	STE_454_178	PF727_STE_AC_00_ER_3_8	7	0	0	0	0
10727	STE_454_182	PF727_STE_AC_22_T1_3_15	0	110	18	53	0
10727	STE_454_183	PF727_STE_AC_22_T1_3_13	0	54	18	0	233
10727	STE_454_190	PF727_STE_AC_32_S1_3_14	0	0	52	42	143
10727	STE_454_197	PF727_STE_AC_00_GE_3_17	0	253	168	136	461
10727	STE_454_301	PF727_STE_AC_32_S1_3_27	0	0	0	37	0

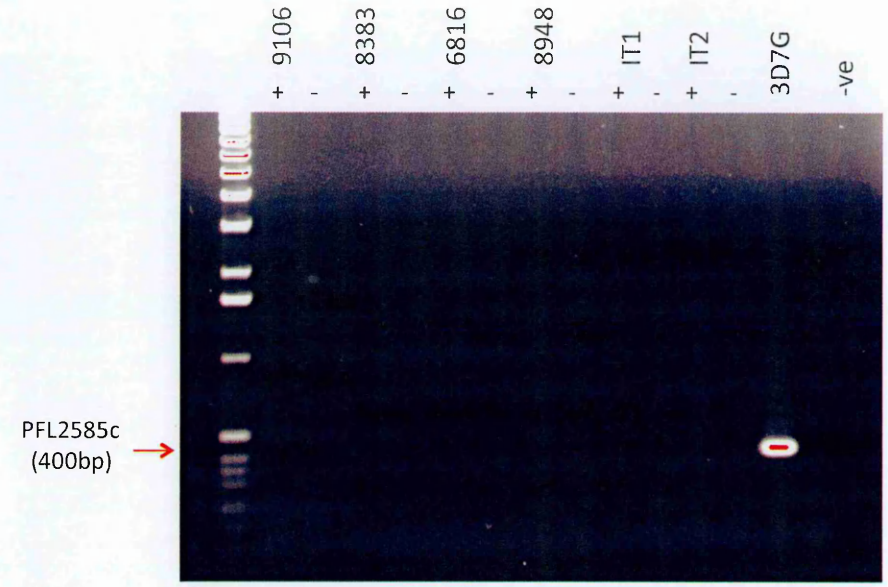
Isolate	Cluster ID	Gene ID	R	T	S	G
10747	STE_454_001	PF747_STE_AC_00_GE_3_1	281	893	1080	63
10747	STE_454_002	PF747_STE_AC_00_GE_3_2	139	75	102	53
10747	STE_454_005	PF747_STE_AC_60_ND_3_6	486	126	225	82
10747	STE_454_007	PF747_STE_AC_40_T1_3_28	0	18	19	26
10747	STE_454_012	PF747_STE_AC_40_T1_3_12	0	8	0	10
10747	STE_454_014	PF747_STE_AC_40_T1_3_10	41	21	39	0
10747	STE_454_016	PF747_STE_AC_60_ND_3_3	163	79	114	175
10747	STE_454_023	PF747_STE_AC_00_GE_3_50	193	78	85	104
10747	STE_454_027	PF747_STE_AC_00_GE_3_42	65	16	5	88
10747	STE_454_028	PF747_STE_AC_00_GE_3_28	105	115	89	396
10747	STE_454_031	PF747_STE_AC_00_GE_3_9	118	88	141	609
10747	STE_454_036	PF747_STE_AC_00_GE_3_5	170	176	164	1309
10747	STE_454_040	PF747_STE_AC_00_GE_3_15	0	6	0	6
10747	STE_454_042	PF747_STE_AC_40_T1_3_18	16	21	20	19
10747	STE_454_048	PF747_STE_AC_00_GE_3_51	80	130	282	226
10747	STE_454_054	PF747_STE_AC_00_GE_3_45	12	60	14	47
10747	STE_454_058	PF747_STE_AC_40_T1_3_27	29	44	32	33
10747	STE_454_063	PF747_STE_AC_00_GE_3_12	124	72	80	24
10747	STE_454_076	PF747_STE_AC_20_R1_3_48	72	0	0	0
10747	STE_454_077	PF747_STE_AC_60_ND_3_10	7	0	39	31
10747	STE_454_079	PF747_STE_AC_60_ND_3_17	32	18	30	15
10747	STE_454_083	PF747_STE_AC_60_ND_3_29	201	40	63	92
10747	STE_454_125	PF747_STE_AC_40_T1_3_5	94	23	18	14
10747	STE_454_136	PF747_STE_AC_20_R1_3_20	47	12	16	28
10747	STE_454_146	PF747_STE_AC_40_T1_3_20	124	68	55	124
10747	STE_454_154	PF747_STE_AC_40_T1_3_7	180	116	136	171
10747	STE_454_158	PF747_STE_AC_00_GE_3_21	152	70	180	126
10747	STE_454_162	PF747_STE_AC_00_GE_3_19	225	318	126	267
10747	STE_454_163	PF747_STE_AC_40_T1_3_17	367	231	262	571
10747	STE_454_172	PF747_STE_AC_40_T1_3_4	15	20	11	35
10747	STE_454_174	PF747_STE_AC_20_R1_3_2	62	106	58	49
10747	STE_454_175	PF747_STE_AC_00_GE_3_6	69	51	62	105
10747	STE_454_177	PF747_STE_AC_20_R1_3_7	6	31	74	85
10747	STE_454_184	PF747_STE_AC_20_R1_3_4	200	211	114	134
10747	STE_454_186	PF747_STE_AC_00_GE_3_17	44	39	57	62
10747	STE_454_187	PF747_STE_AC_20_R1_3_10	206	83	149	155
10747	STE_454_189	PF747_STE_AC_00_GE_3_33	186	96	106	108
10747	STE_454_191	PF747_STE_AC_60_ND_3_23	15	20	18	12
10747	STE_454_192	PF747_STE_AC_20_R1_3_16	22	26	40	42
10747	STE_454_194	PF747_STE_AC_40_T1_3_25	132	73	127	50
10747	STE_454_196	PF747_STE_AC_00_GE_3_37	55	45	55	54
10747	STE_454_198	PF747_STE_AC_00_GE_3_32	77	86	136	165
10747	STE_454_201	PF747_STE_AC_00_GE_3_38	115	49	74	99
10747	STE_454_202	PF747_STE_AC_60_ND_3_26	0	31	34	66
10747	STE_454_203	PF747_STE_AC_00_GE_3_18	704	399	611	502
10747	STE_454_204	PF747_STE_AC_00_GE_3_48	23	17	26	56
10747	STE_454_206	PF747_STE_AC_00_GE_3_4	21	10	16	6
10747	STE_454_217	PF747_STE_AC_40_T1_3_16	17	8	0	12
10747	STE_454_221	PF747_STE_AC_00_GE_3_46	0	20	86	60
10747	STE_454_232	PF747_STE_AC_60_ND_3_12	0	0	8	9
10747	STE_454_264	PF747_STE_AC_00_GE_3_23	0	0	0	14
10747	STE_454_290	PF747_STE_AC_20_R1_3_29	9	0	0	0
10747	STE_454_312	PF747_STE_AC_20_R1_3_50	8	0	0	0

Isolate	Cluster ID	Gene ID	R	LRET	G
10814	STE_454_001	PF814_STE_AC_12_RN_3_1	137	618	73
10814	STE_454_002	PF814_STE_AC_00_GE_3_31	0	0	47
10814	STE_454_005	PF814_STE_AC_12_RN_3_13	6	0	0
10814	STE_454_006	PF814_STE_AC_00_GE_3_14	0	0	21
10814	STE_454_007	PF814_STE_AC_12_RN_3_19	94	7	70
10814	STE_454_008	PF814_STE_AC_00_GE_3_35	0	0	13
10814	STE_454_016	PF814_STE_AC_12_RN_3_6	14	42	568
10814	STE_454_017	PF814_STE_AC_00_GE_3_41	0	0	22
10814	STE_454_020	PF814_STE_AC_12_RN_3_21	23	10	15
10814	STE_454_025	PF814_STE_AC_12_RN_3_11	30	13	210
10814	STE_454_035	PF814_STE_AC_00_GE_3_23	47	11	43
10814	STE_454_042	PF814_STE_AC_00_GE_3_18	0	0	11
10814	STE_454_053	PF814_STE_AC_00_GE_3_1	121	0	272
10814	STE_454_057	PF814_STE_AC_12_RN_3_12	185	681	84
10814	STE_454_059	PF814_STE_AC_12_RN_3_10	12	0	0
10814	STE_454_062	PF814_STE_AC_12_RN_3_23	6	0	9
10814	STE_454_063	PF814_STE_AC_00_GE_3_15	5	0	16
10814	STE_454_066	PF814_STE_AC_00_GE_3_12	0	0	9
10814	STE_454_067	PF814_STE_AC_12_RN_3_5	13	7	31
10814	STE_454_073	PF814_STE_AC_12_RN_3_40	38	0	258
10814	STE_454_086	PF814_STE_AC_12_RN_3_4	38	6	95
10814	STE_454_087	PF814_STE_AC_20_PR_3_19	0	5	150
10814	STE_454_088	PF814_STE_AC_00_GE_3_11	0	0	7
10814	STE_454_091	PF814_STE_AC_00_GE_3_13	6	0	7
10814	STE_454_095	PF814_STE_AC_00_GE_3_26	0	0	41
10814	STE_454_102	PF814_STE_AC_00_GE_3_21	0	0	10
10814	STE_454_109	PF814_STE_AC_12_RN_3_24	36	0	8
10814	STE_454_123	PF814_STE_AC_00_GE_3_32	45	0	37
10814	STE_454_125	PF814_STE_AC_00_GE_3_2	22	0	97
10814	STE_454_127	PF814_STE_AC_12_RN_3_2	8	0	46
10814	STE_454_145	PF814_STE_AC_12_RN_3_18	10	0	66
10814	STE_454_193	PF814_STE_AC_12_RN_3_17	32	0	88
10814	STE_454_199	PF814_STE_AC_20_PR_3_14	25	26	202
10814	STE_454_210	PF814_STE_AC_20_PR_3_3	31	9	109
10814	STE_454_211	PF814_STE_AC_00_GE_3_24	23	7	7
10814	STE_454_213	PF814_STE_AC_20_PR_3_5	22	13	167
10814	STE_454_215	PF814_STE_AC_12_RN_3_14	24	24	20
10814	STE_454_216	PF814_STE_AC_00_GE_3_29	0	0	39
10814	STE_454_218	PF814_STE_AC_20_PR_3_10	92	15	117
10814	STE_454_219	PF814_STE_AC_20_PR_3_12	30	17	150
10814	STE_454_220	PF814_STE_AC_12_RN_3_20	71	8	98
10814	STE_454_224	PF814_STE_AC_12_RN_3_37	11	11	0
10814	STE_454_225	PF814_STE_AC_20_PR_3_18	120	0	279
10814	STE_454_229	PF814_STE_AC_12_RN_3_7	25	0	50
10814	STE_454_230	PF814_STE_AC_12_RN_3_9	27	0	61
10814	STE_454_233	PF814_STE_AC_20_PR_3_22	0	7	0
10814	STE_454_239	PF814_STE_AC_00_GE_3_50	0	0	4
10814	STE_454_242	PF814_STE_AC_00_GE_3_30	8	0	14
10814	STE_454_243	PF814_STE_AC_00_GE_3_33	0	11	78
10814	STE_454_244	PF814_STE_AC_00_GE_3_37	9	0	46
10814	STE_454_247	PF814_STE_AC_00_GE_3_3	0	0	197
10814	STE_454_250	PF814_STE_AC_00_GE_3_8	0	0	5
10814	STE_454_256	PF814_STE_AC_00_GE_3_7	0	0	60
10814	STE_454_276	PF814_STE_AC_12_RN_3_22	19	0	0
10814	STE_454_288	PF814_STE_AC_12_RN_3_31	8	0	0
10814	STE_454_291	PF814_STE_AC_00_GE_3_45	0	0	4

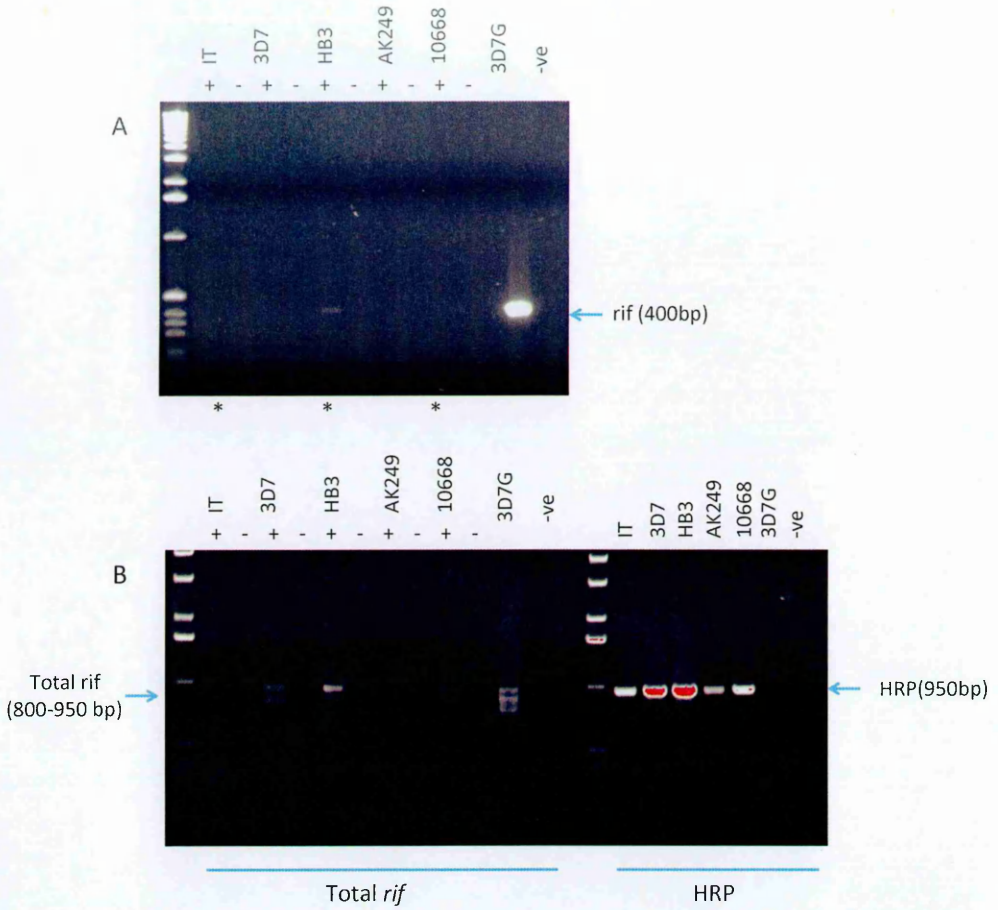


Appendix Figure 9.1. *Agarose gel electrophoresis images of RT-PCR products for PFL2585c (A) and HRP2 (B) from SA075 cDNA at 4 stages of the IDC, rings (1), late ring to early trophozoites (2), trophozoites (3) and schizonts (4), mixed trophozoites cDNA from IT cultured in albumax (IT^a), in pooled serum (IT^p), and in an unrelated IT sample (IT^c), and two different trophozoite samples from a Kilifi isolate 8383 (K1 and K2). G1: It gDNA, G2: 8383 gDNA. G3: SA075 R+ gDNA. –ve: no template control. L: 1 kb plus DNA ladder. (+) and (-) indicate samples with and without the reverse transcriptase enzyme. Arrowheads in (A) indicate samples with incomplete gDNA digestion by the reverse transcriptase enzyme. * HRP PCR on gDNA resulted in many non-specific bands due to suboptimal PCR conditions.*

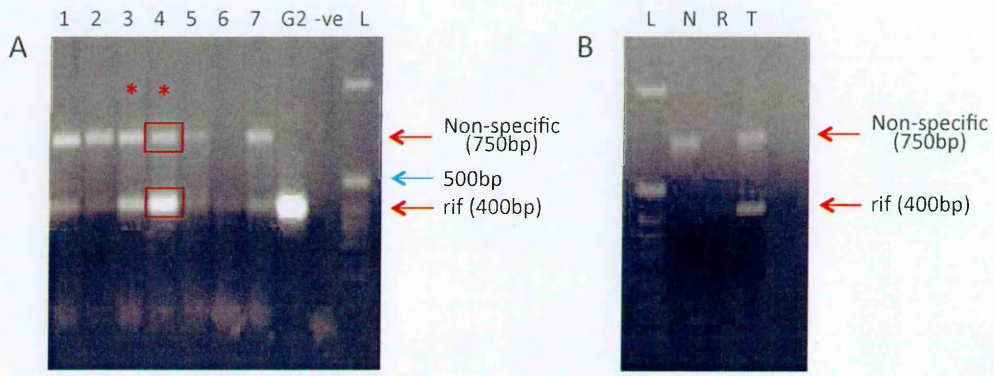
The cDNA used was of good quality as confirmed by HRP PCR. HRP PCR on gDNA did not give a clear band but rather a laddering of bands. This was probably due to suboptimal PCR cobditions, although changing the PCR conditions in a gradient PCR did not result in clear bands from gDNA.



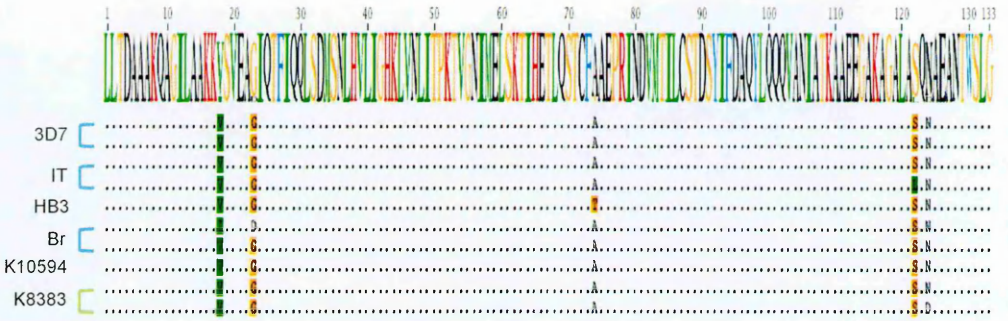
Appendix Figure 9.2. *PFL2585c expression in trophozoites in culture-adapted Kilifi isolates and in laboratory lines, and PCR positive control from 3D7 gDNA. A 400bp band corresponding to PFL2585c HVL was detected in 8383 and the two IT trophozoite fractions sampled at different cycles.*



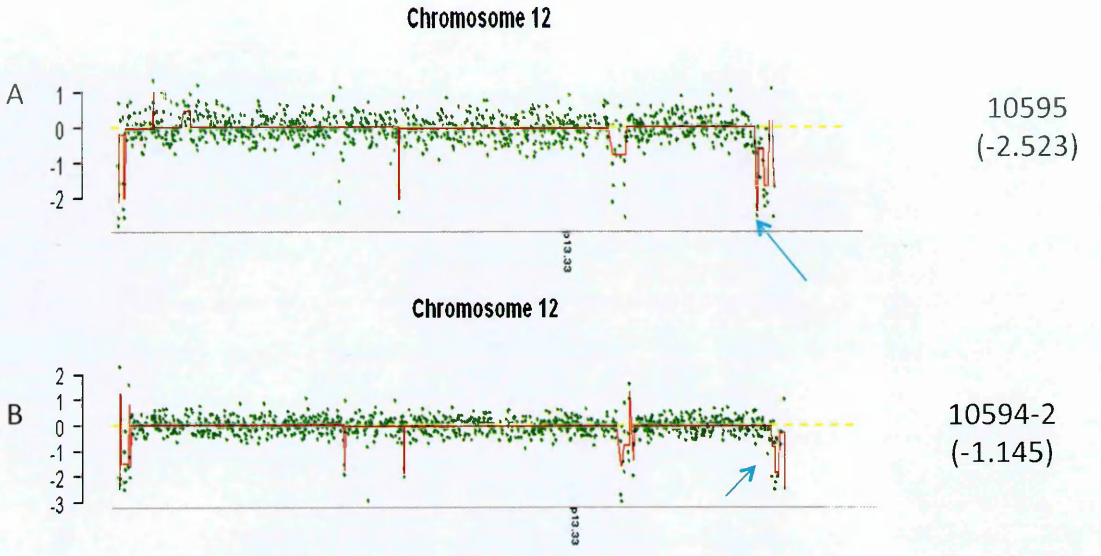
Appendix Figure 9.3. *PFL2585c, total rif and HRP expression in trophozoites from laboratory and culture-adapted Kilifi isolates. (A) PFL2585c amplicons from laboratory lines, IT, 3D7, HB3, and culture-adapted patient isolates AK249 and 10668 from mixed trophozoite stages. The 400pb band corresponding to PFL2585c HVL amplicon was detected faintly in IT, HB3 and in 10668 (*). (B) rif and HRP amplicons for the same samples above. Rif amplicons ranging from 800-950bp were detected in all isolates except in AK249. HRP amplicons were detected in all isolates but not in gDNA. The faint HRP band in AK249 may suggest that the RNA sample used for RT-PCR may have been of poor quality hence the reason why no rif amplicons could be detected.*



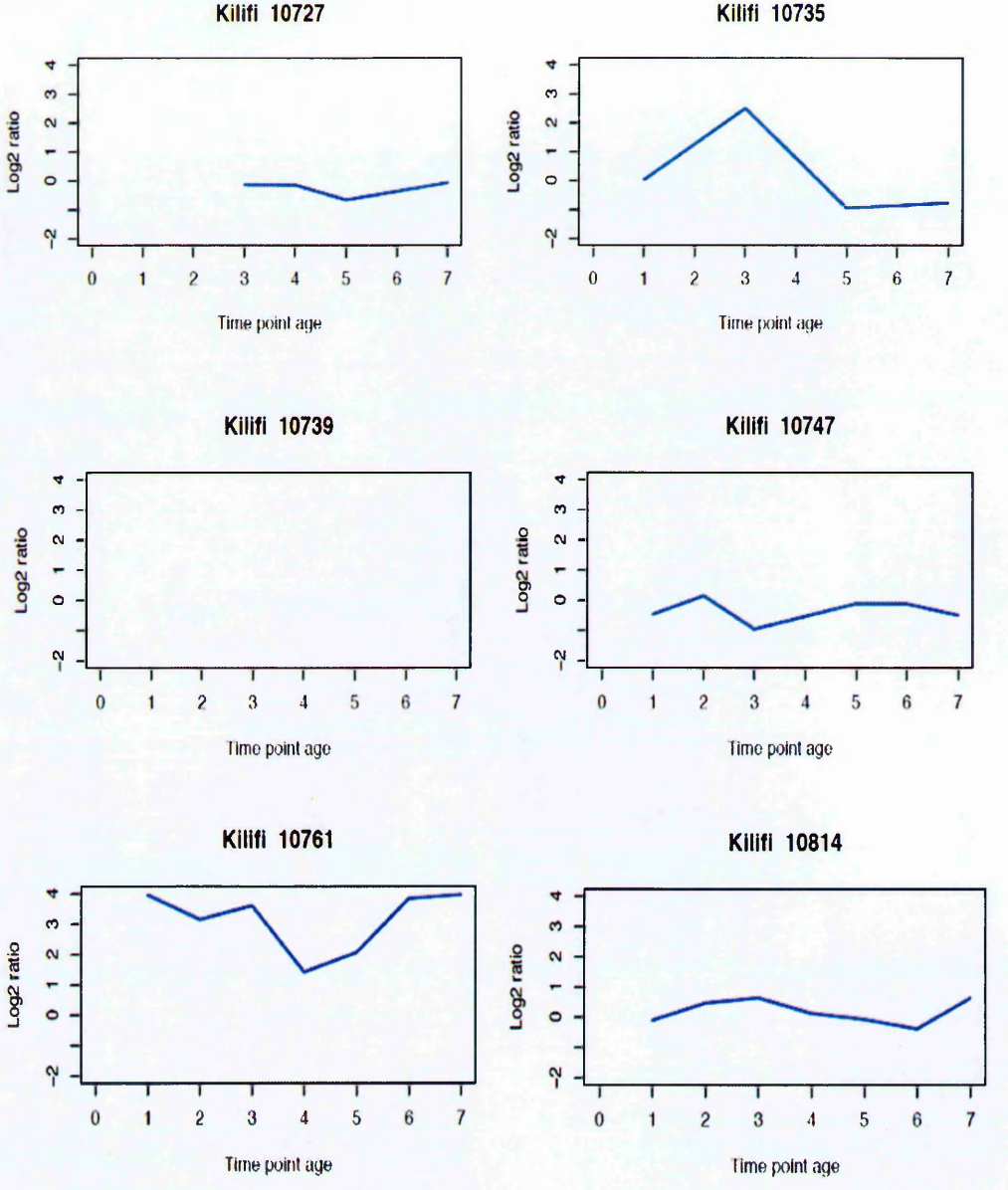
Appendix Figure 9.4. *Cloning PFL2585c RT-PCR amplicons from 10594. (A) Amplification products from PFL2585c specific RT-PCR from a patient isolate 10595 at 7 time points ex vivo. Asterisks (*) indicate amplicons that were purified for cloning by either total PCR cleanup using QIAquick PCR Purification Kit (QIAgen), or gel extraction and purification using the QIAquick Gel Extraction Kit (QIAgen). Red boxes denote bands that were gel extracted. (B) Purified amplicons for cloning. N: The non-PFL2585c upper band of 750bp. R: PFL2585c expected band of 400bp. T: Total PCR purification. L: 1 kb plus DNA ladder.*



Appendix Figure 9.5. *ClustalW alignment of 10 PFL2585c HVL/V2 sequences from 3D7, IT, and HB3 genomes, and 4 wild isolates, 2 from Brazil (Albrecht et al., 2006) and 2 from Kilifi, K10594 and K8383. The 3D7, IT, K8383 and K10594 were obtained by cloning PFL2585c HVL amplicons and capillary sequencing. Mean pairwise identity = 99.2%.*



Appendix Figure 9.6. *Plots of signal (log2ratio) across chromosome 12 in field isolates 10595 (A) and 10594 (B) compared to 3D7. Regions with the red line above or below zero are possible CNVs or highly variable regions (such as VSAs at chromosomal ends). The arrows point to position of the PFL2585c loci. In 10595 (A) the gene lies within a deletion (log2ratio=-2.523). Images courtesy of Mackinnon (Mackinnon et al., unpublished).*



Appendix Figure 9.7. *Plots of log2 ratios for PFL2585c expression from Kilifi isolates at 10 hourly time points of the ex vivo parasite IDC. Images courtesy of Mackinnon (Mackinnon et al., unpublished).*

REFERENCES

ABDEL-LATIF, M. S., CABRERA, G., KOHLER, C., KREMSNER, P. G. & LUTY, A. J. 2004. Antibodies to rifin: a component of naturally acquired responses to Plasmodium falciparum variant surface antigens on infected erythrocytes. *Am J Trop Med Hyg,* 71, 179-86.

ABDEL-LATIF, M. S., DIETZ, K., ISSIFOU, S., KREMSNER, P. G. & KLINKERT, M. Q. 2003. Antibodies to Plasmodium falciparum rifin proteins are associated with rapid parasite clearance and asymptomatic infections. *Infect Immun,* 71, 6229-33.

ABDEL-LATIF, M. S., KHATTAB, A., LINDENTHAL, C., KREMSNER, P. G. & KLINKERT, M. Q. 2002. Recognition of variant Rifin antigens by human antibodies induced during natural Plasmodium falciparum infections. *Infect Immun,* 70, 7013-21.

ABDI, A. I., FEGAN, G., MUTHUI, M., KIRAGU, E., MUSYOKI, J. N., OPIYO, M., MARSH, K., WARIMWE, G. M. & BULL, P. C. 2014. Plasmodium falciparum antigenic variation: relationships between widespread endothelial activation, parasite PfEMP1 expression and severe malaria. *BMC Infect Dis,* 14, 170.

ALBRECHT, L., MERINO, E. F., HOFFMANN, E. H., FERREIRA, M. U., DE MATTOS FERREIRA, R. G., OSAKABE, A. L., DALLA MARTHA, R. C., RAMHARTER, M., DURHAM, A. M., FERREIRA, J. E., DEL PORTILLO, H. A. & WUNDERLICH, G. 2006. Extense variant gene family repertoire overlap in Western Amazon Plasmodium falciparum isolates. *Mol Biochem Parasitol,* 150, 157-65.

AMIN, A. A., ZUROVAC, D., KANGWANA, B. B., GREENFIELD, J., OTIENO, D. N., AKHWALE, W. S. & SNOW, R. W. 2007. The challenges of changing national malaria drug policy to artemisinin-based combinations in Kenya. *Malar J,* 6, 72.

AMINO, R., THIBERGE, S., MARTIN, B., CELLI, S., SHORTE, S., FRISCHKNECHT, F. & MENARD, R. 2006. Quantitative imaging of Plasmodium transmission from mosquito to mammal. *Nat Med,* 12, 220-4.

ANDERS, S. & HUBER, W. 2010. Differential expression analysis for sequence count data. *Genome Biol,* 11, R106.

ASSEFA, S., KEANE, T. M., OTTO, T. D., NEWBOLD, C. & BERRIMAN, M. 2009. ABACAS: algorithm-based automatic contiguation of assembled sequences. *Bioinformatics Vol. 25 No. 15,* 1968-1969.

AVRIL, M., BRAZIER, A. J., MELCHER, M., SAMPATH, S. & SMITH, J. D. 2013. DC8 and DC13 var genes associated with severe malaria bind avidly to diverse endothelial cells. *PLoS Pathog,* 9, e1003430.

AVRIL, M., TRIPATHI, A. K., BRAZIER, A. J., ANDISI, C., JANES, J. H., SOMA, V. L., SULLIVAN, D. J., JR., BULL, P. C., STINS, M. F. & SMITH, J. D. 2012. A restricted subset of var genes mediates adherence of Plasmodium falciparum-infected erythrocytes to brain endothelial cells. *Proc Natl Acad Sci U S A,* 109, E1782-90.

BACHMANN, A., ESSER, C., PETTER, M., PREDEHL, S., VON KALCKREUTH, V., SCHMIEDEL, S., BRUCHHAUS, I. & TANNICH, E. 2009. Absence of erythrocyte sequestration and lack of multicopy gene family expression in

Plasmodium falciparum from a splenectomized malaria patient. *PLoS One,* 4, e7459.
BACHMANN, A., PETTER, M., TILLY, A. K., BILLER, L., ULICZKA, K. A., DUFFY, M. F., TANNICH, E. & BRUCHHAUS, I. 2012. Temporal expression and localization patterns of variant surface antigens in clinical Plasmodium falciparum isolates during erythrocyte schizogony. *PLoS One,* 7, e49540.
BACHMANN, A., SCHOLZ, J. A., JANSSEN, M., KLINKERT, M. Q., TANNICH, E., BRUCHHAUS, I. & PETTER, M. 2015. A comparative study of the localization and membrane topology of members of the RIFIN, STEVOR and PfMC-2TM protein families in Plasmodium falciparum-infected erythrocytes. *Malar J,* 14, 274.
BARNWELL, J. W., HOWARD, R. J., COON, H. G. & MILLER, L. H. 1983. Splenic requirement for antigenic variation and expression of the variant antigen on the erythrocyte membrane in cloned Plasmodium knowlesi malaria. *Infect Immun,* 40, 985-94.
BARNWELL, J. W., HOWARD, R. J. & MILLER, L. H. 1982. Altered expression of Plasmodium knowlesi variant antigen on the erythrocyte membrane in splenectomized rhesus monkeys. *J Immunol,* 128, 224-6.
BARUCH, D. I., PASLOSKE, B. L., SINGH, H. B., BI, X., MA, X. C., FELDMAN, M., TARASCHI, T. F. & HOWARD, R. J. 1995. Cloning the P. falciparum gene encoding PfEMP1, a malarial variant antigen and adherence receptor on the surface of parasitized human erythrocytes. *Cell,* 82, 77-87.
BAUM, J., GILBERGER, T. W., FRISCHKNECHT, F. & MEISSNER, M. 2008. Host-cell invasion by malaria parasites: insights from Plasmodium and Toxoplasma. *Trends Parasitol,* 24, 557-63.
BECK, J. R., MURALIDHARAN, V., OKSMAN, A. & GOLDBERG, D. E. 2014. PTEX component HSP101 mediates export of diverse malaria effectors into host erythrocytes. *Nature,* 511, 592-5.
BEESON, J. G., REEDER, J. C., ROGERSON, S. J. & BROWN, G. V. 2001. Parasite adhesion and immune evasion in placental malaria. *Trends Parasitol,* 17, 331-7.
BEJON, P., MWACHARO, J., KAI, O., MWANGI, T., MILLIGAN, P., TODRYK, S., KEATING, S., LANG, T., LOWE, B., GIKONYO, C., MOLYNEUX, C., FEGAN, G., GILBERT, S. C., PESHU, N., MARSH, K. & HILL, A. V. 2006. A phase 2b randomised trial of the candidate malaria vaccines FP9 ME-TRAP and MVA ME-TRAP among children in Kenya. *PLoS Clin Trials,* 1, e29.
BEJON, P., WILLIAMS, T. N., LILJANDER, A., NOOR, A. M., WAMBUA, J., OGADA, E., OLOTU, A., OSIER, F. H., HAY, S. I., FARNERT, A. & MARSH, K. 2010. Stable and unstable malaria hotspots in longitudinal cohort studies in Kenya. *PLoS Med,* 7, e1000304.
BEJON, P., WILLIAMS, T. N., NYUNDO, C., HAY, S. I., BENZ, D., GETHING, P. W., OTIENDE, M., PESHU, J., BASHRAHEIL, M., GREENHOUSE, B., BOUSEMA, T., BAUNI, E., MARSH, K., SMITH, D. L. & BORRMANN, S. 2014. A micro-epidemiological analysis of febrile malaria in Coastal Kenya showing hotspots within hotspots. *Elife,* 3, e02130.
BERGER, S. S., TURNER, L., WANG, C. W., PETERSEN, J. E., KRAFT, M., LUSINGU, J. P., MMBANDO, B., MARQUARD, A. M., BENGTSSON, D. B.,

HVIID, L., NIELSEN, M. A., THEANDER, T. G. & LAVSTSEN, T. 2013. Plasmodium falciparum expressing domain cassette 5 type PfEMP1 (DC5-PfEMP1) bind PECAM1. *PLoS One,* 8, e69117.

BERNABEU, M., LOPEZ, F. J., FERRER, M., MARTIN-JAULAR, L., RAZANAME, A., CORRADIN, G., MAIER, A. G., DEL PORTILLO, H. A. & FERNANDEZ-BECERRA, C. 2012. Functional analysis of Plasmodium vivax VIR proteins reveals different subcellular localizations and cytoadherence to the ICAM-1 endothelial receptor. *Cell Microbiol,* 14, 386-400.

BIGGS, B. A., GOOZE, L., WYCHERLEY, K., WOLLISH, W., SOUTHWELL, B., LEECH, J. H. & BROWN, G. V. 1991. Antigenic variation in Plasmodium falciparum. *Proc Natl Acad Sci U S A,* 88, 9171-4.

BLYTHE, J. E., NIANG, M., MARSH, K., HOLDER, A. A., LANGHORNE, J. & PREISER, P. R. 2009. Characterization of the repertoire diversity of the Plasmodium falciparum stevor multigene family in laboratory and field isolates. *Malar J,* 8, 140.

BLYTHE, J. E., SURENTHERAN, T. & PREISER, P. R. 2004. STEVOR--a multifunctional protein? *Mol Biochem Parasitol,* 134, 11-5.

BLYTHE, J. E., YAM, X. Y., KUSS, C., BOZDECH, Z., HOLDER, A. A., MARSH, K., LANGHORNE, J. & PREISER, P. R. 2008. Plasmodium falciparum STEVOR proteins are highly expressed in patient isolates and located in the surface membranes of infected red blood cells and the apical tips of merozoites. *Infect Immun,* 76, 3329-36.

BODDEY, J. A., CARVALHO, T. G., HODDER, A. N., SARGEANT, T. J., SLEEBS, B. E., MARAPANA, D., LOPATICKI, S., NEBL, T. & COWMAN, A. F. 2013. Role of plasmepsin V in export of diverse protein families from the Plasmodium falciparum exportome. *Traffic,* 14, 532-50.

BOETZER, M., HENKEL, C., JANSEN, H., BUTLER, D. & PIROVANO, W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics 27(4),* 578-579.

BOUSEMA, T. & DRAKELEY, C. 2011. Epidemiology and infectivity of Plasmodium falciparum and Plasmodium vivax gametocytes in relation to malaria control and elimination. *Clin Microbiol Rev,* 24, 377-410.

BOZDECH, Z., LLINAS, M., PULLIAM, B. L., WONG, E. D., ZHU, J. & DERISI, J. L. 2003. The transcriptome of the intraerythrocytic developmental cycle of Plasmodium falciparum. *PLoS Biol,* 1, E5.

BROWN, I. N., BROWN, K. N. & HILLS, L. A. 1968. Immunity to malaria: the antibody response to antigenic variation by Plasmodium knowlesi. *Immunology,* 14, 127-38.

BROWN, K. N. & BROWN, I. N. 1965. Immunity to malaria: antigenic variation in chronic infections of Plasmodium knowlesi. *Nature,* 208, 1286-8.

BULL, P. C., BERRIMAN, M., KYES, S., QUAIL, M. A., HALL, N., KORTOK, M. M., MARSH, K. & NEWBOLD, C. I. 2005. Plasmodium falciparum variant surface antigen expression patterns during malaria. *PLoS Pathog,* 1, e26.

BULL, P. C., BUCKEE, C. O., KYES, S., KORTOK, M. M., THATHY, V., GUYAH, B., STOUTE, J. A., NEWBOLD, C. I. & MARSH, K. 2008. Plasmodium falciparum antigenic variation. Mapping mosaic var gene sequences onto a

network of shared, highly polymorphic sequence blocks. *Mol Microbiol,* 68, 1519-34.

BULL, P. C., KORTOK, M., KAI, O., NDUNGU, F., ROSS, A., LOWE, B. S., NEWBOLD, C. I. & MARSH, K. 2000. Plasmodium falciparum-infected erythrocytes: agglutination by diverse Kenyan plasma is associated with severe disease and young host age. *J Infect Dis,* 182, 252-9.

BULL, P. C., LOWE, B. S., KORTOK, M. & MARSH, K. 1999. Antibody recognition of Plasmodium falciparum erythrocyte surface antigens in Kenya: evidence for rare and prevalent variants. *Infect Immun,* 67, 733-9.

BULL, P. C., LOWE, B. S., KORTOK, M., MOLYNEUX, C. S., NEWBOLD, C. I. & MARSH, K. 1998. Parasite antigens on the infected red cell surface are targets for naturally acquired immunity to malaria. *Nat Med,* 4, 358-60.

BULL, P. C. & MARSH, K. 2002. The role of antibodies to Plasmodium falciparum-infected-erythrocyte surface antigens in naturally acquired immunity to malaria. *Trends Microbiol,* 10, 55-8.

BULTRINI, E., BRICK, K., MUKHERJEE, S., ZHANG, Y., SILVESTRINI, F., ALANO, P. & PIZZI, E. 2009. Revisiting the Plasmodium falciparum RIFIN family: from comparative genomics to 3D-model prediction. *BMC Genomics,* 10, 445.

CABRAL, F. J. & WUNDERLICH, G. 2009. Transcriptional memory and switching in the Plasmodium falciparumrif gene family. *Mol Biochem Parasitol,* 168, 186-90.

CARLTON, J. M., ADAMS, J. H., SILVA, J. C., BIDWELL, S. L., LORENZI, H., CALER, E., CRABTREE, J., ANGIUOLI, S. V., MERINO, E. F., AMEDEO, P., CHENG, Q., COULSON, R. M., CRABB, B. S., DEL PORTILLO, H. A., ESSIEN, K., FELDBLYUM, T. V., FERNANDEZ-BECERRA, C., GILSON, P. R., GUEYE, A. H., GUO, X., KANG'A, S., KOOIJ, T. W., KORSINCZKY, M., MEYER, E. V., NENE, V., PAULSEN, I., WHITE, O., RALPH, S. A., REN, Q., SARGEANT, T. J., SALZBERG, S. L., STOECKERT, C. J., SULLIVAN, S. A., YAMAMOTO, M. M., HOFFMAN, S. L., WORTMAN, J. R., GARDNER, M. J., GALINSKI, M. R., BARNWELL, J. W. & FRASER-LIGGETT, C. M. 2008. Comparative genomics of the neglected human malaria parasite Plasmodium vivax. *Nature,* 455, 757-63.

CARLTON, J. M., ANGIUOLI, S. V., SUH, B. B., KOOIJ, T. W., PERTEA, M., SILVA, J. C., ERMOLAEVA, M. D., ALLEN, J. E., SELENGUT, J. D., KOO, H. L., PETERSON, J. D., POP, M., KOSACK, D. S., SHUMWAY, M. F., BIDWELL, S. L., SHALLOM, S. J., VAN AKEN, S. E., RIEDMULLER, S. B., FELDBLYUM, T. V., CHO, J. K., QUACKENBUSH, J., SEDEGAH, M., SHOAIBI, A., CUMMINGS, L. M., FLORENS, L., YATES, J. R., RAINE, J. D., SINDEN, R. E., HARRIS, M. A., CUNNINGHAM, D. A., PREISER, P. R., BERGMAN, L. W., VAIDYA, A. B., VAN LIN, L. H., JANSE, C. J., WATERS, A. P., SMITH, H. O., WHITE, O. R., SALZBERG, S. L., VENTER, J. C., FRASER, C. M., HOFFMAN, S. L., GARDNER, M. J. & CARUCCI, D. J. 2002. Genome sequence and comparative analysis of the model rodent malaria parasite Plasmodium yoelii yoelii. *Nature,* 419, 512-9.

CARVALHO, B. O., LOPES, S. C., NOGUEIRA, P. A., ORLANDI, P. P., BARGIERI, D. Y., BLANCO, Y. C., MAMONI, R., LEITE, J. A., RODRIGUES, M. M.,

SOARES, I. S., OLIVEIRA, T. R., WUNDERLICH, G., LACERDA, M. V., DEL PORTILLO, H. A., ARAUJO, M. O., RUSSELL, B., SUWANARUSK, R., SNOUNOU, G., RENIA, L. & COSTA, F. T. 2010. On the cytoadhesion of Plasmodium vivax-infected erythrocytes. *J Infect Dis,* 202, 638-47.

CARVER, T., HARRIS, S. R., OTTO, T. D., BERRIMAN, M., PARKHILL, J. & MCQUILLAN, J. A. 2013. BamView: visualizing and interpretation of next-generation sequencing read alignments. *Brief Bioinform,* 14, 203-12.

CEESAY, S. J., CASALS-PASCUAL, C., ERSKINE, J., ANYA, S. E., DUAH, N. O., FULFORD, A. J., SESAY, S. S., ABUBAKAR, I., DUNYO, S., SEY, O., PALMER, A., FOFANA, M., CORRAH, T., BOJANG, K. A., WHITTLE, H. C., GREENWOOD, B. M. & CONWAY, D. J. 2008. Changes in malaria indices between 1999 and 2007 in The Gambia: a retrospective analysis. *Lancet,* 372, 1545-54.

CHAN, J. A., HOWELL, K. B., REILING, L., ATAIDE, R., MACKINTOSH, C. L., FOWKES, F. J., PETTER, M., CHESSON, J. M., LANGER, C., WARIMWE, G. M., DUFFY, M. F., ROGERSON, S. J., BULL, P. C., COWMAN, A. F., MARSH, K. & BEESON, J. G. 2012. Targets of antibodies against Plasmodium falciparum-infected erythrocytes in malaria immunity. *J Clin Invest,* 122, 3227-38.

CHEN, Q., FERNANDEZ, V., SUNDSTROM, A., SCHLICHTHERLE, M., DATTA, S., HAGBLOM, P. & WAHLGREN, M. 1998. Developmental selection of var gene expression in Plasmodium falciparum. *Nature,* 394, 392-5.

CHENG, Q., CLOONAN, N., FISCHER, K., THOMPSON, J., WAINE, G., LANZER, M. & SAUL, A. 1998. stevor and rif are Plasmodium falciparum multicopy gene families which potentially encode variant antigens. *Mol Biochem Parasitol,* 97, 161-76.

CHOOKAJORN, T., DZIKOWSKI, R., FRANK, M., LI, F., JIWANI, A. Z., HARTL, D. L. & DEITSCH, K. W. 2007. Epigenetic memory at malaria virulence genes. *Proc Natl Acad Sci U S A,* 104, 899-902.

CLAESSENS, A., ADAMS, Y., GHUMRA, A., LINDERGARD, G., BUCHAN, C. C., ANDISI, C., BULL, P. C., MOK, S., GUPTA, A. P., WANG, C. W., TURNER, L., ARMAN, M., RAZA, A., BOZDECH, Z. & ROWE, J. A. 2012. A subset of group A-like var genes encodes the malaria parasite ligands for binding to human brain endothelial cells. *Proc Natl Acad Sci U S A,* 109, E1772-81.

CLAESSENS, A., GHUMRA, A., GUPTA, A. P., MOK, S., BOZDECH, Z. & ROWE, J. A. 2011. Design of a variant surface antigen-supplemented microarray chip for whole transcriptome analysis of multiple Plasmodium falciparum cytoadherent strains, and identification of strain-transcendent rif and stevor genes. *Malar J,* 10, 180.

COHEN, S., MC, G. I. & CARRINGTON, S. 1961. Gamma-globulin and acquired immunity to human malaria. *Nature,* 192, 733-7.

CUNNINGHAM, D., FONAGER, J., JARRA, W., CARRET, C., PREISER, P. & LANGHORNE, J. 2009. Rapid changes in transcription profiles of the Plasmodium yoelii yir multigene family in clonal populations: lack of epigenetic memory? *PLoS One,* 4, e4285.

CUNNINGHAM, D., LAWTON, J., JARRA, W., PREISER, P. & LANGHORNE, J. 2010. The pir multigene family of Plasmodium: antigenic variation and beyond. *Mol Biochem Parasitol,* 170, 65-73.

DAILY, J. P., LE ROCH, K. G., SARR, O., NDIAYE, D., LUKENS, A., ZHOU, Y., NDIR, O., MBOUP, S., SULTAN, A., WINZELER, E. A. & WIRTH, D. F. 2005. In vivo transcriptome of Plasmodium falciparum reveals overexpression of transcripts that encode surface proteins. *J Infect Dis,* 191, 1196-203.

DE KONING-WARD, T. F., GILSON, P. R., BODDEY, J. A., RUG, M., SMITH, B. J., PAPENFUSS, A. T., SANDERS, P. R., LUNDIE, R. J., MAIER, A. G., COWMAN, A. F. & CRABB, B. S. 2009. A newly discovered protein export machine in malaria parasites. *Nature,* 459, 945-9.

DEL PORTILLO, H. A., FERNANDEZ-BECERRA, C., BOWMAN, S., OLIVER, K., PREUSS, M., SANCHEZ, C. P., SCHNEIDER, N. K., VILLALOBOS, J. M., RAJANDREAM, M. A., HARRIS, D., PEREIRA DA SILVA, L. H., BARRELL, B. & LANZER, M. 2001. A superfamily of variant genes encoded in the subtelomeric region of Plasmodium vivax. *Nature,* 410, 839-42.

DEL PORTILLO, H. A., LANZER, M., RODRIGUEZ-MALAGA, S., ZAVALA, F. & FERNANDEZ-BECERRA, C. 2004. Variant genes and the spleen in Plasmodium vivax malaria. *Int J Parasitol,* 34, 1547-54.

DODOO, D., STAALSOE, T., GIHA, H., KURTZHALS, J. A., AKANMORI, B. D., KORAM, K., DUNYO, S., NKRUMAH, F. K., HVIID, L. & THEANDER, T. G. 2001. Antibodies to variant antigens on the surfaces of infected erythrocytes are associated with protection from malaria in Ghanaian children. *Infect Immun,* 69, 3713-8.

DOOLAN, D. L., DOBANO, C. & BAIRD, J. K. 2009. Acquired immunity to malaria. *Clin Microbiol Rev,* 22, 13-36, Table of Contents.

DRAKELEY, C. J., BOUSEMA, J. T., AKIM, N. I., TEELEN, K., ROEFFEN, W., LENSEN, A. H., BOLMER, M., ELING, W. & SAUERWEIN, R. W. 2006. Transmission-reducing immunity is inversely related to age in Plasmodium falciparum gametocyte carriers. *Parasite Immunol,* 28, 185-90.

DUFFY, P. E. & FRIED, M. 2003a. Antibodies that inhibit Plasmodium falciparum adhesion to chondroitin sulfate A are associated with increased birth weight and the gestational age of newborns. *Infect Immun,* 71, 6620-3.

DUFFY, P. E. & FRIED, M. 2003b. Plasmodium falciparum adhesion in the placenta. *Curr Opin Microbiol,* 6, 371-6.

DYER, M. & DAY, K. P. 2000. Commitment to gametocytogenesis in Plasmodium falciparum. *Parasitol Today,* 16, 102-7.

DZIKOWSKI, R. & DEITSCH, K. W. 2008. Active transcription is required for maintenance of epigenetic memory in the malaria parasite Plasmodium falciparum. *J Mol Biol,* 382, 288-97.

DZIKOWSKI, R., FRANK, M. & DEITSCH, K. 2006a. Mutually exclusive expression of virulence genes by malaria parasites is regulated independently of antigen production. *PLoS Pathog,* 2, e22.

DZIKOWSKI, R., LI, F., AMULIC, B., EISBERG, A., FRANK, M., PATEL, S., WELLEMS, T. E. & DEITSCH, K. W. 2007. Mechanisms underlying mutually

exclusive expression of virulence genes by malaria parasites. *EMBO Rep,* 8, 959-65.
DZIKOWSKI, R., TEMPLETON, T. J. & DEITSCH, K. 2006b. Variant antigen gene expression in malaria. *Cell Microbiol,* 8, 1371-81.
EATON, M. D. 1938. The Agglutination of Plasmodium Knowlesi by Immune Serum. *J Exp Med,* 67, 857-70.
EDGAR, R. C. 2004a. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics,* 5, 113.
EDGAR, R. C. 2004b. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res,* 32, 1792-7.
EDOZIEN 1962. Adult cord-blood gamma-globulin and immunity to malaria in Nigerians. *The Lancet,* 1962, 951-955.
ELSWORTH, B., MATTHEWS, K., NIE, C. Q., KALANON, M., CHARNAUD, S. C., SANDERS, P. R., CHISHOLM, S. A., COUNIHAN, N. A., SHAW, P. J., PINO, P., CHAN, J. A., AZEVEDO, M. F., ROGERSON, S. J., BEESON, J. G., CRABB, B. S., GILSON, P. R. & DE KONING-WARD, T. F. 2014. PTEX is an essential nexus for protein export in malaria parasites. *Nature,* 511, 587-91.
FERNANDEZ, V., HOMMEL, M., CHEN, Q., HAGBLOM, P. & WAHLGREN, M. 1999. Small, clonally variant antigens expressed on the surface of the Plasmodium falciparum-infected erythrocyte are encoded by the rif gene family and are the target of human immune responses. *J Exp Med,* 190, 1393-404.
FERNANDEZ-BECERRA, C., PEIN, O., DE OLIVEIRA, T. R., YAMAMOTO, M. M., CASSOLA, A. C., ROCHA, C., SOARES, I. S., DE BRAGANCA PEREIRA, C. A. & DEL PORTILLO, H. A. 2005. Variant proteins of Plasmodium vivax are not clonally expressed in natural infections. *Mol Microbiol,* 58, 648-58.
FISCHER, K., CHAVCHICH, M., HUESTIS, R., WILSON, D. W., KEMP, D. J. & SAUL, A. 2003. Ten families of variant genes encoded in subtelomeric regions of multiple chromosomes of Plasmodium chabaudi, a malaria species that undergoes antigenic variation in the laboratory mouse. *Mol Microbiol,* 48, 1209-23.
FLORENS, L., LIU, X., WANG, Y., YANG, S., SCHWARTZ, O., PEGLAR, M., CARUCCI, D. J., YATES, J. R., 3RD & WUB, Y. 2004. Proteomics approach reveals novel proteins on the surface of malaria-infected erythrocytes. *Mol Biochem Parasitol,* 135, 1-11.
FLORENS, L., WASHBURN, M. P., RAINE, J. D., ANTHONY, R. M., GRAINGER, M., HAYNES, J. D., MOCH, J. K., MUSTER, N., SACCI, J. B., TABB, D. L., WITNEY, A. A., WOLTERS, D., WU, Y., GARDNER, M. J., HOLDER, A. A., SINDEN, R. E., YATES, J. R. & CARUCCI, D. J. 2002. A proteomic view of the Plasmodium falciparum life cycle. *Nature,* 419, 520-6.
FRANK, M., DZIKOWSKI, R., COSTANTINI, D., AMULIC, B., BERDOUGO, E. & DEITSCH, K. 2006. Strict pairing of var promoters and introns is required for var gene silencing in the malaria parasite Plasmodium falciparum. *J Biol Chem,* 281, 9942-52.
FRIED, M. & DUFFY, P. E. 1998. Maternal malaria and parasite adhesion. *J Mol Med,* 76, 162-71.
FRISCHKNECHT, F., BALDACCI, P., MARTIN, B., ZIMMER, C., THIBERGE, S., OLIVO-MARIN, J. C., SHORTE, S. L. & MENARD, R. 2004. Imaging

movement of malaria parasites during transmission by Anopheles mosquitoes. *Cell Microbiol,* 6, 687-94.

FTP://FTP.SANGER.AC.UK/PUB/PROJECT/PATHOGENS/PLASMODIUM/FALCIPARUM/HB3/ 2013.

GANNOUN-ZAKI, L., JOST, A., MU, J., DEITSCH, K. W. & WELLEMS, T. E. 2005. A silenced Plasmodium falciparum var promoter can be activated in vivo through spontaneous deletion of a silencing element in the intron. *Eukaryot Cell,* 4, 490-2.

GARCIA, J. E., PUENTES, A., CURTIDOR, H., VERA, R., RODRIGUEZ, L., VALBUENA, J., LOPEZ, R., OCAMPO, M., CORTES, J., VANEGAS, M., ROSAS, J., REYES, C. & PATARROYO, M. E. 2005. Peptides from the Plasmodium falciparum STEVOR putative protein bind with high affinity to normal human red blood cells. *Peptides,* 26, 1133-43.

GARDNER, J. P., PINCHES, R. A., ROBERTS, D. J. & NEWBOLD, C. I. 1996. Variant antigens and endothelial receptor adhesion in Plasmodium falciparum. *Proc Natl Acad Sci U S A,* 93, 3503-8.

GARDNER, M. J., HALL, N., FUNG, E., WHITE, O., BERRIMAN, M., HYMAN, R. W., CARLTON, J. M., PAIN, A., NELSON, K. E., BOWMAN, S., PAULSEN, I. T., JAMES, K., EISEN, J. A., RUTHERFORD, K., SALZBERG, S. L., CRAIG, A., KYES, S., CHAN, M. S., NENE, V., SHALLOM, S. J., SUH, B., PETERSON, J., ANGIUOLI, S., PERTEA, M., ALLEN, J., SELENGUT, J., HAFT, D., MATHER, M. W., VAIDYA, A. B., MARTIN, D. M., FAIRLAMB, A. H., FRAUNHOLZ, M. J., ROOS, D. S., RALPH, S. A., MCFADDEN, G. I., CUMMINGS, L. M., SUBRAMANIAN, G. M., MUNGALL, C., VENTER, J. C., CARUCCI, D. J., HOFFMAN, S. L., NEWBOLD, C., DAVIS, R. W., FRASER, C. M. & BARRELL, B. 2002. Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature,* 419, 498-511.

GARDNER, M. J., TETTELIN, H., CARUCCI, D. J., CUMMINGS, L. M., ARAVIND, L., KOONIN, E. V., SHALLOM, S., MASON, T., YU, K., FUJII, C., PEDERSON, J., SHEN, K., JING, J., ASTON, C., LAI, Z., SCHWARTZ, D. C., PERTEA, M., SALZBERG, S., ZHOU, L., SUTTON, G. G., CLAYTON, R., WHITE, O., SMITH, H. O., FRASER, C. M., ADAMS, M. D., VENTER, J. C. & HOFFMAN, S. L. 1998. Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum. *Science,* 282, 1126-32.

GIHA, H. A., STAALSOE, T., DODOO, D., ROPER, C., SATTI, G. M., ARNOT, D. E., HVIID, L. & THEANDER, T. G. 2000. Antibodies to variable Plasmodium falciparum-infected erythrocyte surface antigens are associated with protection from novel malaria infections. *Immunol Lett,* 71, 117-26.

GINSBURG, H., KRUGLIAK, M., EIDELMAN, O. & CABANTCHIK, Z. I. 1983. New permeability pathways induced in membranes of Plasmodium falciparum infected erythrocytes. *Mol Biochem Parasitol,* 8, 177-90.

GINSBURG, H. & STEIN, W. D. 1987. New permeability pathways induced by the malarial parasite in the membrane of its host erythrocyte: potential routes for targeting of drugs into infected cells. *Biosci Rep,* 7, 455-63.

GINSBURG, H. & STEIN, W. D. 2004. The new permeability pathways induced by the malaria parasite in the membrane of the infected erythrocyte: comparison of results using different experimental techniques. *J Membr Biol,* 197, 113-34.

GOEL, S., PALMKVIST, M., MOLL, K., JOANNIN, N., LARA, P., R, R. A., MORADI, N., OJEMALM, K., WESTMAN, M., ANGELETTI, D., KJELLIN, H., LEHTIO, J., BLIXT, O., IDESTROM, L., GAHMBERG, C. G., STORRY, J. R., HULT, A. K., OLSSON, M. L., VON HEIJNE, G., NILSSON, I. & WAHLGREN, M. 2015. RIFINs are adhesins implicated in severe Plasmodium falciparum malaria. *Nat Med,* 21, 314-7.

GRAEWE, S., RANKIN, K. E., LEHMANN, C., DESCHERMEIER, C., HECHT, L., FROEHLKE, U., STANWAY, R. R. & HEUSSLER, V. 2011. Hostile takeover by Plasmodium: reorganization of parasite and host cell membranes during liver stage egress. *PLoS Pathog,* 7, e1002224.

GRURING, C., HEIBER, A., KRUSE, F., FLEMMING, S., FRANCI, G., COLOMBO, S. F., FASANA, E., SCHOELER, H., BORGESE, N., STUNNENBERG, H. G., PRZYBORSKI, J. M., GILBERGER, T. W. & SPIELMANN, T. 2012. Uncovering common principles in protein export of malaria parasites. *Cell Host Microbe,* 12, 717-29.

GUINDON, S., DELSUC, F., DUFAYARD, J. F. & GASCUEL, O. 2009. Estimating maximum likelihood phylogenies with PhyML. *Methods Mol Biol,* 537, 113-37.

GUINDON, S. & GASCUEL, O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol,* 52, 696-704.

GUINDON, S., LETHIEC, F., DUROUX, P. & GASCUEL, O. 2005. PHYML Online--a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res,* 33, W557-9.

GUPTA, S., SNOW, R. W., DONNELLY, C. A., MARSH, K. & NEWBOLD, C. 1999. Immunity to non-cerebral severe malaria is acquired after one or two infections. *Nat Med,* 5, 340-3.

HAEGGSTROM, M., A, V. O. N. E., KIRONDE, F., FERNANDEZ, V. & WAHLGREN, M. 2007. Characterization of Maurer's clefts in Plasmodium falciparum-infected erythrocytes. *Am J Trop Med Hyg,* 76, 27-32.

HAEGGSTROM, M., KIRONDE, F., BERZINS, K., CHEN, Q., WAHLGREN, M. & FERNANDEZ, V. 2004. Common trafficking pathway for variant antigens destined for the surface of the Plasmodium falciparum-infected erythrocyte. *Mol Biochem Parasitol,* 133, 1-14.

HELMBY, H., CAVELIER, L., PETTERSSON, U. & WAHLGREN, M. 1993. Rosetting Plasmodium falciparum-infected erythrocytes express unique strain-specific antigens on their surface. *Infect Immun,* 61, 284-8.

HILLER, N. L., BHATTACHARJEE, S., VAN OOIJ, C., LIOLIOS, K., HARRISON, T., LOPEZ-ESTRANO, C. & HALDAR, K. 2004. A host-targeting signal in virulence proteins reveals a secretome in malarial infection. *Science,* 306, 1934-7.

HOMMEL, M., DAVID, P. H. & OLIGINO, L. D. 1983. Surface alterations of erythrocytes in Plasmodium falciparum malaria. Antigenic variation, antigenic diversity, and the role of the spleen. *J Exp Med,* 157, 1137-48.

HORROCKS, P., PINCHES, R., CHRISTODOULOU, Z., KYES, S. A. & NEWBOLD, C. I. 2004. Variable var transition rates underlie antigenic variation in malaria. *Proc Natl Acad Sci U S A,* 101, 11129-34.

HOWARD, R. J., BARNWELL, J. W. & KAO, V. 1983. Antigenic variation of Plasmodium knowlesi malaria: identification of the variant antigen on infected erythrocytes. *Proc Natl Acad Sci U S A,* 80, 4129-33.
HOWITT, C. A., WILINSKI, D., LLINAS, M., TEMPLETON, T. J., DZIKOWSKI, R. & DEITSCH, K. W. 2009. Clonally variant gene families in Plasmodium falciparum share a common activation factor. *Mol Microbiol,* 73, 1171-85.
HTTP://WWW.ATGC-MONTPELLIER.FR.
HTTP://WWW.BROAD.MIT.EDU. Broad Institute of Harvard and MIT.
HTTP://WWW.EBI.AC.UK/TOOLS/MSA/MUSCLE/.
HTTP://WWW.GENEDB.ORG/GENEDB/MALARIA/. Wellcome Trust Sanger Institute.
HTTP://WWW.PLASMODB.ORG/PLASMO/HOME.JSP.
HTTP://WWW.SANGER.AC.UK/PROJECTS/P_FALCIPARUM. Wellcome Trust Sanger Institute.
HUANG, X. & MADAN, A. 1999. CAP3: A DNA sequence assembly program. *Genome Res,* 9, 868-77.
HUANG, Y., NIU, B., GAO, Y., FU, L. & LI, W. 2010. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics,* 26, 680-2.
JANSSEN, C. S., PHILLIPS, R. S., TURNER, C. M. & BARRETT, M. P. 2004. Plasmodium interspersed repeats: the major multigene superfamily of malaria parasites. *Nucleic Acids Res,* 32, 5712-20.
JEMMELY, N. Y., NIANG, M. & PREISER, P. R. 2010. Small variant surface antigens and Plasmodium evasion of immunity. *Future Microbiol,* 5, 663-82.
JENSEN, A. T., MAGISTRADO, P., SHARP, S., JOERGENSEN, L., LAVSTSEN, T., CHIUCCHIUINI, A., SALANTI, A., VESTERGAARD, L. S., LUSINGU, J. P., HERMSEN, R., SAUERWEIN, R., CHRISTENSEN, J., NIELSEN, M. A., HVIID, L., SUTHERLAND, C., STAALSOE, T. & THEANDER, T. G. 2004. Plasmodium falciparum associated with severe childhood malaria preferentially expresses PfEMP1 encoded by group A var genes. *J Exp Med,* 199, 1179-90.
JIN, Y., KEBAIER, C. & VANDERBERG, J. 2007. Direct microscopic quantification of dynamics of Plasmodium berghei sporozoite transmission from mosquitoes to mice. *Infect Immun,* 75, 5532-9.
JOANNIN, N., ABHIMAN, S., SONNHAMMER, E. L. & WAHLGREN, M. 2008. Sub-grouping and sub-functionalization of the RIFIN multi-copy protein family. *BMC Genomics,* 9, 19.
JOANNIN, N., KALLBERG, Y., WAHLGREN, M. & PERSSON, B. 2011. RSpred, a set of Hidden Markov Models to detect and classify the RIFIN and STEVOR proteins of Plasmodium falciparum. *BMC Genomics,* 12, 119.
KAVIRATNE, M., KHAN, S. M., JARRA, W. & PREISER, P. R. 2002. Small variant STEVOR antigen is uniquely located within Maurer's clefts in Plasmodium falciparum-infected red blood cells. *Eukaryot Cell,* 1, 926-35.
KHATTAB, A., BONOW, I., SCHREIBER, N., PETTER, M., SCHMETZ, C. & KLINKERT, M. Q. 2008. Plasmodium falciparum variant STEVOR antigens are expressed in merozoites and possibly associated with erythrocyte invasion. *Malar J,* 7, 137.

KHATTAB, A. & MERI, S. 2011. Exposure of the Plasmodium falciparum clonally variant STEVOR proteins on the merozoite surface. *Malar J,* 10, 58.

KIDGELL, C., VOLKMAN, S. K., DAILY, J., BOREVITZ, J. O., PLOUFFE, D., ZHOU, Y., JOHNSON, J. R., LE ROCH, K., SARR, O., NDIR, O., MBOUP, S., BATALOV, S., WIRTH, D. F. & WINZELER, E. A. 2006. A systematic map of genetic variation in Plasmodium falciparum. *PLoS Pathog,* 2, e57.

KLEINSCHMIDT, I., SCHWABE, C., BENAVENTE, L., TORREZ, M., RIDL, F. C., SEGURA, J. L., EHMER, P. & NCHAMA, G. N. 2009. Marked increase in child survival after four years of intensive malaria control. *Am J Trop Med Hyg,* 80, 882-8.

KOZAREWA, I., NING, Z., QUAIL, M. A., SANDERS, M. J., BERRIMAN, M. & TURNER, D. J. 2009. Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. *Nat Methods,* 6, 291-5.

KRAEMER, S. M. & SMITH, J. D. 2003. Evidence for the importance of genetic structuring to the structural and functional specialization of the Plasmodium falciparum var gene family. *Mol Microbiol,* 50, 1527-38.

KYES, S., PINCHES, R. & NEWBOLD, C. 2000. A simple RNA analysis method shows var and rif multigene family expression patterns in Plasmodium falciparum. *Mol Biochem Parasitol,* 105, 311-5.

KYES, S. A., ROWE, J. A., KRIEK, N. & NEWBOLD, C. I. 1999. Rifins: a second family of clonally variant proteins expressed on the surface of red cells infected with Plasmodium falciparum. *Proc Natl Acad Sci U S A,* 96, 9333-8.

LANGHORNE, J., NDUNGU, F. M., SPONAAS, A. M. & MARSH, K. 2008. Immunity to malaria: more questions than answers. *Nat Immunol,* 9, 725-32.

LANGRETH, S. G. & REESE, R. T. 1979. Antigenicity of the infected-erythrocyte and merozoite surfaces in Falciparum malaria. *J Exp Med,* 150, 1241-54.

LARKIN, M. A., BLACKSHIELDS, G., BROWN, N. P., CHENNA, R., MCGETTIGAN, P. A., MCWILLIAM, H., VALENTIN, F., WALLACE, I. M., WILM, A., LOPEZ, R., THOMPSON, J. D., GIBSON, T. J. & HIGGINS, D. G. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics,* 23, 2947-8.

LAVAZEC, C., SANYAL, S. & TEMPLETON, T. J. 2006. Hypervariability within the Rifin, Stevor and Pfmc-2TM superfamilies in Plasmodium falciparum. *Nucleic Acids Res,* 34, 6696-707.

LAVAZEC, C., SANYAL, S. & TEMPLETON, T. J. 2007. Expression switching in the stevor and Pfmc-2TM superfamilies in Plasmodium falciparum. *Mol Microbiol,* 64, 1621-34.

LAVSTSEN, T., SALANTI, A., JENSEN, A. T., ARNOT, D. E. & THEANDER, T. G. 2003. Sub-grouping of Plasmodium falciparum 3D7 var genes based on sequence analysis of coding and non-coding regions. *Malar J,* 2, 27.

LAVSTSEN, T., TURNER, L., SAGUTI, F., MAGISTRADO, P., RASK, T. S., JESPERSEN, J. S., WANG, C. W., BERGER, S. S., BARAKA, V., MARQUARD, A. M., SEGUIN-ORLANDO, A., WILLERSLEV, E., GILBERT, M. T., LUSINGU, J. & THEANDER, T. G. 2012. Plasmodium falciparum erythrocyte membrane protein 1 domain cassettes 8 and 13 are associated with severe malaria in children. *Proc Natl Acad Sci U S A,* 109, E1791-800.

LAWTON, J., BRUGAT, T., YAN, Y. X., REID, A. J., BOHME, U., OTTO, T. D., PAIN, A., JACKSON, A., BERRIMAN, M., CUNNINGHAM, D., PREISER, P. & LANGHORNE, J. 2012. Characterization and gene expression analysis of the cir multi-gene family of Plasmodium chabaudi chabaudi (AS). *BMC Genomics,* 13, 125.

LE ROCH, K. G., ZHOU, Y., BLAIR, P. L., GRAINGER, M., MOCH, J. K., HAYNES, J. D., DE LA VEGA, P., HOLDER, A. A., BATALOV, S., CARUCCI, D. J. & WINZELER, E. A. 2003. Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science,* 301, 1503-8.

LEECH, J. H., BARNWELL, J. W., MILLER, L. H. & HOWARD, R. J. 1984. Identification of a strain-specific malarial antigen exposed on the surface of Plasmodium falciparum-infected erythrocytes. *J Exp Med,* 159, 1567-75.

LEMIEUX, J. E., GOMEZ-ESCOBAR, N., FELLER, A., CARRET, C., AMAMBUA-NGWA, A., PINCHES, R., DAY, F., KYES, S. A., CONWAY, D. J., HOLMES, C. C. & NEWBOLD, C. I. 2009. Statistical estimation of cell-cycle progression and lineage commitment in Plasmodium falciparum reveals a homogeneous pattern of transcription in ex vivo culture. *Proc Natl Acad Sci U S A,* 106, 7559-64.

LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G., DURBIN, R. & GENOME PROJECT DATA PROCESSING, S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25, 2078-9.

LI, W. & GODZIK, A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics,* 22, 1658-9.

LIMPAIBOON, T., TAYLOR, D. W., JONES, G., GEYSEN, H. M. & SAUL, A. 1990. Characterization of a Plasmodium falciparum epitope recognized by a monoclonal antibody with broad isolate and species specificity. *Southeast Asian J Trop Med Public Health,* 21, 388-96.

LLINAS, M., BOZDECH, Z., WONG, E. D., ADAI, A. T. & DERISI, J. L. 2006. Comparative whole genome transcriptome analysis of three Plasmodium falciparum strains. *Nucleic Acids Res,* 34, 1166-73.

LOGAN-KLUMPLER, F. J., DE SILVA, N., BOEHME, U., ROGERS, M. B., VELARDE, G., MCQUILLAN, J. A., CARVER, T., ASLETT, M., OLSEN, C., SUBRAMANIAN, S., PHAN, I., FARRIS, C., MITRA, S., RAMASAMY, G., WANG, H., TIVEY, A., JACKSON, A., HOUSTON, R., PARKHILL, J., HOLDEN, M., HARB, O. S., BRUNK, B. P., MYLER, P. J., ROOS, D., CARRINGTON, M., SMITH, D. F., HERTZ-FOWLER, C. & BERRIMAN, M. 2012. GeneDB--an annotation database for pathogens. *Nucleic Acids Res,* 40, D98-108.

LOPEZ, F. J., BERNABEU, M., FERNANDEZ-BECERRA, C. & DEL PORTILLO, H. A. 2013. A new computational approach redefines the subtelomeric vir superfamily of Plasmodium vivax. *BMC Genomics,* 14, 8.

MAIER, A. G., COOKE, B. M., COWMAN, A. F. & TILLEY, L. 2009. Malaria parasite proteins that remodel the host erythrocyte. *Nat Rev Microbiol,* 7, 341-54.

MAIER, A. G., RUG, M., O'NEILL, M. T., BROWN, M., CHAKRAVORTY, S., SZESTAK, T., CHESSON, J., WU, Y., HUGHES, K., COPPEL, R. L.,

NEWBOLD, C., BEESON, J. G., CRAIG, A., CRABB, B. S. & COWMAN, A. F. 2008. Exported proteins required for virulence and rigidity of Plasmodium falciparum-infected human erythrocytes. *Cell,* 134, 48-61.

MARGULIES, M., EGHOLM, M., ALTMAN, W. E., ATTIYA, S., BADER, J. S., BEMBEN, L. A., BERKA, J., BRAVERMAN, M. S., CHEN, Y. J., CHEN, Z., DEWELL, S. B., DU, L., FIERRO, J. M., GOMES, X. V., GODWIN, B. C., HE, W., HELGESEN, S., HO, C. H., IRZYK, G. P., JANDO, S. C., ALENQUER, M. L., JARVIE, T. P., JIRAGE, K. B., KIM, J. B., KNIGHT, J. R., LANZA, J. R., LEAMON, J. H., LEFKOWITZ, S. M., LEI, M., LI, J., LOHMAN, K. L., LU, H., MAKHIJANI, V. B., MCDADE, K. E., MCKENNA, M. P., MYERS, E. W., NICKERSON, E., NOBILE, J. R., PLANT, R., PUC, B. P., RONAN, M. T., ROTH, G. T., SARKIS, G. J., SIMONS, J. F., SIMPSON, J. W., SRINIVASAN, M., TARTARO, K. R., TOMASZ, A., VOGT, K. A., VOLKMER, G. A., WANG, S. H., WANG, Y., WEINER, M. P., YU, P., BEGLEY, R. F. & ROTHBERG, J. M. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature,* 437, 376-80.

MARSH, K. 1992. Malaria--a neglected disease? *Parasitology,* 104 Suppl, S53-69.

MARSH, K., FORSTER, D., WARUIRU, C., MWANGI, I., WINSTANLEY, M., MARSH, V., NEWTON, C., WINSTANLEY, P., WARN, P., PESHU, N. & ET AL. 1995. Indicators of life-threatening malaria in African children. *N Engl J Med,* 332, 1399-404.

MARSH, K. & HOWARD, R. J. 1986. Antigens induced on erythrocytes by P. falciparum: expression of diverse and conserved determinants. *Science,* 231, 150-3.

MARSH, K. & KINYANJUI, S. 2006. Immune effector mechanisms in malaria. *Parasite Immunol,* 28, 51-60.

MARSH, K., OTOO, L., HAYES, R. J., CARSON, D. C. & GREENWOOD, B. M. 1989. Antibodies to blood stage antigens of Plasmodium falciparum in rural Gambians and their relation to protection against infection. *Trans R Soc Trop Med Hyg,* 83, 293-303.

MARTI, M., GOOD, R. T., RUG, M., KNUEPFER, E. & COWMAN, A. F. 2004. Targeting malaria virulence and remodeling proteins to the host erythrocyte. *Science,* 306, 1930-3.

MBOGO, C. M., MWANGANGI, J. M., NZOVU, J., GU, W., YAN, G., GUNTER, J. T., SWALM, C., KEATING, J., REGENS, J. L., SHILILU, J. I., GITHURE, J. I. & BEIER, J. C. 2003. Spatial and temporal heterogeneity of Anopheles mosquitoes and Plasmodium falciparum transmission along the Kenyan coast. *Am J Trop Med Hyg,* 68, 734-42.

MCBRIDE, J. S., NEWBOLD, C. I. & ANAND, R. 1985. Polymorphism of a high molecular weight schizont antigen of the human malaria parasite Plasmodium falciparum. *J Exp Med,* 161, 160-80.

MCROBERT, L., PREISER, P., SHARP, S., JARRA, W., KAVIRATNE, M., TAYLOR, M. C., RENIA, L. & SUTHERLAND, C. J. 2004. Distinct trafficking and localization of STEVOR proteins in three stages of the Plasmodium falciparum life cycle. *Infect Immun,* 72, 6597-602.

MEDICA, D. L. & SINNIS, P. 2005. Quantitative dynamics of Plasmodium yoelii sporozoite transmission by infected anopheline mosquitoes. *Infect Immun,* 73, 4363-9.
MENARD, R. 2001. Gliding motility and cell invasion by Apicomplexa: insights from the Plasmodium sporozoite. *Cell Microbiol,* 3, 63-73.
MENARD, R. 2005. Medicine: knockout malaria vaccine? *Nature,* 433, 113-4.
MERCEREAU-PUIJALON, O., GUILLOTTE, M. & VIGAN-WOMAS, I. 2008. Rosetting in Plasmodium falciparum: a cytoadherence phenotype with multiple actors. *Transfus Clin Biol,* 15, 62-71.
MOLYNEUX, M. E., TAYLOR, T. E., WIRIMA, J. J. & BORGSTEIN, A. 1989. Clinical features and prognostic indicators in paediatric cerebral malaria: a study of 131 comatose Malawian children. *Q J Med,* 71, 441-59.
MOTA, M. M., HAFALLA, J. C. & RODRIGUEZ, A. 2002. Migration through host cells activates Plasmodium sporozoites for infection. *Nat Med,* 8, 1318-22.
MOTA, M. M., PRADEL, G., VANDERBERG, J. P., HAFALLA, J. C., FREVERT, U., NUSSENZWEIG, R. S., NUSSENZWEIG, V. & RODRIGUEZ, A. 2001. Migration of Plasmodium sporozoites through cells before infection. *Science,* 291, 141-4.
MOTA, M. M. & RODRIGUEZ, A. 2004. Migration through host cells: the first steps of Plasmodium sporozoites in the mammalian host. *Cell Microbiol,* 6, 1113-8.
MOXON, C. A., WASSMER, S. C., MILNER, D. A., JR., CHISALA, N. V., TAYLOR, T. E., SEYDEL, K. B., MOLYNEUX, M. E., FARAGHER, B., ESMON, C. T., DOWNEY, C., TOH, C. H., CRAIG, A. G. & HEYDERMAN, R. S. 2013. Loss of endothelial protein C receptors links coagulation and inflammation to parasite sequestration in cerebral malaria in African children. *Blood.*
MWAKALINGA, S. B., WANG, C. W., BENGTSSON, D. C., TURNER, L., DINKO, B., LUSINGU, J. P., ARNOT, D. E., SUTHERLAND, C. J., THEANDER, T. G. & LAVSTSEN, T. 2012. Expression of a type B RIFIN in Plasmodium falciparum merozoites and gametes. *Malar J,* 11, 429.
MWANGANGI, J. M., MBOGO, C. M., ORINDI, B. O., MUTURI, E. J., MIDEGA, J. T., NZOVU, J., GATAKAA, H., GITHURE, J., BORGEMEISTER, C., KEATING, J. & BEIER, J. C. 2013. Shifts in malaria vector species composition and transmission dynamics along the Kenyan coast over the past 20 years. *Malar J,* 12, 13.
MWANGI, T. W., ROSS, A., SNOW, R. W. & MARSH, K. 2005. Case definitions of clinical malaria under different transmission conditions in Kilifi District, Kenya. *J Infect Dis,* 191, 1932-9.
NEWBOLD, C., CRAIG, A., KYES, S., ROWE, A., FERNANDEZ-REYES, D. & FAGAN, T. 1999. Cytoadherence, pathogenesis and the infected red cell surface in Plasmodium falciparum. *Int J Parasitol,* 29, 927-37.
NEWBOLD, C. I. 1999. Antigenic variation in Plasmodium falciparum: mechanisms and consequences. *Curr Opin Microbiol,* 2, 420-5.
NEWBOLD, C. I., PINCHES, R., ROBERTS, D. J. & MARSH, K. 1992. Plasmodium falciparum: the human agglutinating antibody response to the infected red cell surface is predominantly variant specific. *Exp Parasitol,* 75, 281-92.

NEWTON, C. R., CHOKWE, T., SCHELLENBERG, J. A., WINSTANLEY, P. A., FORSTER, D., PESHU, N., KIRKHAM, F. J. & MARSH, K. 1997. Coma scales for children with severe falciparum malaria. *Trans R Soc Trop Med Hyg,* 91, 161-5.

NIANG, M., BEI, A. K., MADNANI, K. G., PELLY, S., DANKWA, S., KANJEE, U., GUNALAN, K., AMALADOSS, A., YEO, K. P., BOB, N. S., MALLERET, B., DURAISINGH, M. T. & PREISER, P. R. 2014. STEVOR is a Plasmodium falciparum erythrocyte binding protein that mediates merozoite invasion and rosetting. *Cell Host Microbe,* 16, 81-93.

NIANG, M., YAN YAM, X. & PREISER, P. R. 2009. The Plasmodium falciparum STEVOR multigene family mediates antigenic variation of the infected erythrocyte. *PLoS Pathog,* 5, e1000307.

NIELSEN, M. A., STAALSOE, T., KURTZHALS, J. A., GOKA, B. Q., DODOO, D., ALIFRANGIS, M., THEANDER, T. G., AKANMORI, B. D. & HVIID, L. 2002. Plasmodium falciparum variant surface antigen expression varies between isolates causing severe and nonsevere malaria and is modified by acquired immunity. *J Immunol,* 168, 3444-50.

O'MEARA, W. P., BEJON, P., MWANGI, T. W., OKIRO, E. A., PESHU, N., SNOW, R. W., NEWTON, C. R. & MARSH, K. 2008. Effect of a fall in malaria transmission on morbidity and mortality in Kilifi, Kenya. *Lancet,* 372, 1555-62.

OKIRO, E. A., HAY, S. I., GIKANDI, P. W., SHARIF, S. K., NOOR, A. M., PESHU, N., MARSH, K. & SNOW, R. W. 2007. The decline in paediatric malaria admissions on the coast of Kenya. *Malar J,* 6, 151.

OTTO, T. D. 2015. From sequence mapping to genome assemblies. *Methods Mol Biol,* 1201, 19-50.

OTTO, T. D., BOHME, U., JACKSON, A. P., HUNT, M., FRANKE-FAYARD, B., HOEIJMAKERS, W. A., RELIGA, A. A., ROBERTSON, L., SANDERS, M., OGUN, S. A., CUNNINGHAM, D., ERHART, A., BILLKER, O., KHAN, S. M., STUNNENBERG, H. G., LANGHORNE, J., HOLDER, A. A., WATERS, A. P., NEWBOLD, C. I., PAIN, A., BERRIMAN, M. & JANSE, C. J. 2014a. A comprehensive evaluation of rodent malaria parasite genomes and gene expression. *BMC Biol,* 12, 86.

OTTO, T. D., DILLON, G. P., DEGRAVE, W. S. & BERRIMAN, M. 2011. RATT: Rapid Annotation Transfer Tool. *Nucleic Acids Research,* 1-7.

OTTO, T. D., RAYNER, J. C., BOHME, U., PAIN, A., SPOTTISWOODE, N., SANDERS, M., QUAIL, M., OLLOMO, B., RENAUD, F., THOMAS, A. W., PRUGNOLLE, F., CONWAY, D. J., NEWBOLD, C. & BERRIMAN, M. 2014b. Genome sequencing of chimpanzee malaria parasites reveals possible pathways of adaptation to human hosts. *Nat Commun,* 5, 4754.

OTTO, T. D., WILINSKI, D., ASSEFA, S., KEANE, T. M., SARRY, L. R., BOHME, U., LEMIEUX, J., BARRELL, B., PAIN, A., BERRIMAN, M., NEWBOLD, C. & LLINAS, M. 2010. New insights into the blood-stage transcriptome of Plasmodium falciparum using RNA-Seq. *Mol Microbiol,* 76, 12-24.

PAIN, A., BOHME, U., BERRY, A. E., MUNGALL, K., FINN, R. D., JACKSON, A. P., MOURIER, T., MISTRY, J., PASINI, E. M., ASLETT, M. A., BALASUBRAMMANIAM, S., BORGWARDT, K., BROOKS, K., CARRET,

C., CARVER, T. J., CHEREVACH, I., CHILLINGWORTH, T., CLARK, T. G., GALINSKI, M. R., HALL, N., HARPER, D., HARRIS, D., HAUSER, H., IVENS, A., JANSSEN, C. S., KEANE, T., LARKE, N., LAPP, S., MARTI, M., MOULE, S., MEYER, I. M., ORMOND, D., PETERS, N., SANDERS, M., SANDERS, S., SARGEANT, T. J., SIMMONDS, M., SMITH, F., SQUARES, R., THURSTON, S., TIVEY, A. R., WALKER, D., WHITE, B., ZUIDERWIJK, E., CHURCHER, C., QUAIL, M. A., COWMAN, A. F., TURNER, C. M., RAJANDREAM, M. A., KOCKEN, C. H., THOMAS, A. W., NEWBOLD, C. I., BARRELL, B. G. & BERRIMAN, M. 2008. The genome of the simian and human malaria parasite Plasmodium knowlesi. *Nature,* 455, 799-803.

PETTER, M., BONOW, I. & KLINKERT, M. Q. 2008. Diverse expression patterns of subgroups of the rif multigene family during Plasmodium falciparum gametocytogenesis. *PLoS ONE,* 3, e3779.

PETTER, M., HAEGGSTROM, M., KHATTAB, A., FERNANDEZ, V., KLINKERT, M. Q. & WAHLGREN, M. 2007. Variant proteins of the Plasmodium falciparum RIFIN family show distinct subcellular localization and developmental expression patterns. *Mol Biochem Parasitol,* 156, 51-61.

PIMENTA, P. F., TOURAY, M. & MILLER, L. 1994. The journey of malaria sporozoites in the mosquito salivary gland. *J Eukaryot Microbiol,* 41, 608-24.

PRADEL, G. 2007. Proteins of the malaria parasite sexual stages: expression, function and potential for transmission blocking strategies. *Parasitology,* 134, 1911-29.

PREISER, P. R. 2002. Parasite packs a punch. *Nat Med,* 8, 1198-9.

PRZYBORSKI, J. M., MILLER, S. K., PFAHLER, J. M., HENRICH, P. P., ROHRBACH, P., CRABB, B. S. & LANZER, M. 2005. Trafficking of STEVOR to the Maurer's clefts in Plasmodium falciparum-infected erythrocytes. *Embo J,* 24, 2306-17.

RASK, T. S., HANSEN, D. A., THEANDER, T. G., GORM PEDERSEN, A. & LAVSTSEN, T. 2010. Plasmodium falciparum erythrocyte membrane protein 1 diversity in seven genomes--divide and conquer. *PLoS Comput Biol,* 6.

REEDER, J. C., COWMAN, A. F., DAVERN, K. M., BEESON, J. G., THOMPSON, J. K., ROGERSON, S. J. & BROWN, G. V. 1999. The adhesion of Plasmodium falciparum-infected erythrocytes to chondroitin sulfate A is mediated by P. falciparum erythrocyte membrane protein 1. *Proc Natl Acad Sci U S A,* 96, 5198-202.

RIBACKE, U., MOK, B. W., WIRTA, V., NORMARK, J., LUNDEBERG, J., KIRONDE, F., EGWANG, T. G., NILSSON, P. & WAHLGREN, M. 2007. Genome wide gene amplifications and deletions in Plasmodium falciparum. *Mol Biochem Parasitol,* 155, 33-44.

RIGLAR, D. T., RICHARD, D., WILSON, D. W., BOYLE, M. J., DEKIWADIA, C., TURNBULL, L., ANGRISANO, F., MARAPANA, D. S., ROGERS, K. L., WHITCHURCH, C. B., BEESON, J. G., COWMAN, A. F., RALPH, S. A. & BAUM, J. 2011. Super-resolution dissection of coordinated events during malaria parasite invasion of the human erythrocyte. *Cell Host Microbe,* 9, 9-20.

RIGLAR, D. T., ROGERS, K. L., HANSSEN, E., TURNBULL, L., BULLEN, H. E., CHARNAUD, S. C., PRZYBORSKI, J., GILSON, P. R., WHITCHURCH, C. B., CRABB, B. S., BAUM, J. & COWMAN, A. F. 2013. Spatial association with

PTEX complexes defines regions for effector export into Plasmodium falciparum-infected erythrocytes. *Nat Commun,* 4, 1415.
ROBERTS, D. J., CRAIG, A. G., BERENDT, A. R., PINCHES, R., NASH, G., MARSH, K. & NEWBOLD, C. I. 1992. Rapid switching to multiple antigenic and adhesive phenotypes in malaria. *Nature,* 357, 689-92.
ROBINSON, B. A., WELCH, T. L. & SMITH, J. D. 2003. Widespread functional specialization of Plasmodium falciparum erythrocyte membrane protein 1 family members to bind CD36 analysed across a parasite genome. *Mol Microbiol,* 47, 1265-78.
ROVIRA-GRAELLS, N., GUPTA, A. P., PLANET, E., CROWLEY, V. M., MOK, S., RIBAS DE POUPLANA, L., PREISER, P. R., BOZDECH, Z. & CORTES, A. 2012. Transcriptional variation in the malaria parasite Plasmodium falciparum. *Genome Res,* 22, 925-38.
ROWE, J. A., MOULDS, J. M., NEWBOLD, C. I. & MILLER, L. H. 1997. P. falciparum rosetting mediated by a parasite-variant erythrocyte membrane protein and complement-receptor 1. *Nature,* 388, 292-5.
ROWE, J. A., SCRAGG, I. G., KWIATKOWSKI, D., FERGUSON, D. J., CARUCCI, D. J. & NEWBOLD, C. I. 1998. Implications of mycoplasma contamination in Plasmodium falciparum cultures and methods for its detection and eradication. *Mol Biochem Parasitol,* 92, 177-80.
RUG, M., CYRKLAFF, M., MIKKONEN, A., LEMGRUBER, L., KUELZER, S., SANCHEZ, C. P., THOMPSON, J., HANSSEN, E., O'NEILL, M., LANGER, C., LANZER, M., FRISCHKNECHT, F., MAIER, A. G. & COWMAN, A. F. 2014. Export of virulence proteins by malaria-infected erythrocytes involves remodeling of host actin cytoskeleton. *Blood,* 124, 3459-68.
RUTHERFORD, K., PARKHILL, J., CROOK, J., HORSNELL, T., RICE, P., RAJANDREAM, M. A. & BARRELL, B. 2000. Artemis: sequence visualization and annotation. *Bioinformatics,* 16, 944-5.
SALANTI, A., STAALSOE, T., LAVSTSEN, T., JENSEN, A. T., SOWA, M. P., ARNOT, D. E., HVIID, L. & THEANDER, T. G. 2003. Selective upregulation of a single distinctly structured var gene in chondroitin sulphate A-adhering Plasmodium falciparum involved in pregnancy-associated malaria. *Mol Microbiol,* 49, 179-91.
SAM-YELLOWE, T. Y., FLORENS, L., JOHNSON, J. R., WANG, T., DRAZBA, J. A., LE ROCH, K. G., ZHOU, Y., BATALOV, S., CARUCCI, D. J., WINZELER, E. A. & YATES, J. R., 3RD 2004. A Plasmodium gene family encoding Maurer's cleft membrane proteins: structural properties and expression profiling. *Genome Res,* 14, 1052-9.
SANCHEZ, C. 2011. Parasitology: The malaria food channel. *Nat Rev Microbiol,* 9, 484.
SANYAL, S., EGEE, S., BOUYER, G., PERROT, S., SAFEUKUI, I., BISCHOFF, E., BUFFET, P., DEITSCH, K. W., MERCEREAU-PUIJALON, O., DAVID, P. H., TEMPLETON, T. J. & LAVAZEC, C. 2012. Plasmodium falciparum STEVOR proteins impact erythrocyte mechanical properties. *Blood,* 119, e1-8.
SCHERF, A., HERNANDEZ-RIVAS, R., BUFFET, P., BOTTIUS, E., BENATAR, C., POUVELLE, B., GYSIN, J. & LANZER, M. 1998. Antigenic variation in malaria: in situ switching, relaxed and mutually exclusive transcription of var

genes during intra-erythrocytic development in Plasmodium falciparum. *Embo J,* 17, 5418-26.

SCHERF, A., LOPEZ-RUBIO, J. J. & RIVIERE, L. 2008. Antigenic variation in Plasmodium falciparum. *Annu Rev Microbiol,* 62, 445-70.

SCHREIBER, N., BRATTIG, N., EVANS, J., TSIRI, A., HORSTMANN, R. D., MAY, J. & KLINKERT, M. Q. 2006. Cerebral malaria is associated with IgG2 and IgG4 antibody responses to recombinant Plasmodium falciparum RIFIN antigen. *Microbes Infect,* 8, 1269-76.

SCHREIBER, N., KHATTAB, A., PETTER, M., MARKS, F., ADJEI, S., KOBBE, R., MAY, J. & KLINKERT, M. Q. 2008. Expression of Plasmodium falciparum 3D7 STEVOR proteins for evaluation of antibody responses following malaria infections in naive infants. *Parasitology,* 135, 155-67.

SHARP, S., LAVSTSEN, T., FIVELMAN, Q. L., SAEED, M., MCROBERT, L., TEMPLETON, T. J., JENSEN, A. T., BAKER, D. A., THEANDER, T. G. & SUTHERLAND, C. J. 2006. Programmed transcription of the var gene family, but not of stevor, in Plasmodium falciparum gametocytes. *Eukaryot Cell,* 5, 1206-14.

SHOCK, J. L., FISCHER, K. F. & DERISI, J. L. 2007. Whole-genome analysis of mRNA decay in Plasmodium falciparum reveals a global lengthening of mRNA half-life during the intra-erythrocytic development cycle. *Genome Biol,* 8, R134.

SHRETTA, R., OMUMBO, J., RAPUODA, B. & SNOW, R. W. 2000. Using evidence to change antimalarial drug policy in Kenya. *Trop Med Int Health,* 5, 755-64.

SINGH, B., KIM SUNG, L., MATUSOP, A., RADHAKRISHNAN, A., SHAMSUL, S. S., COX-SINGH, J., THOMAS, A. & CONWAY, D. J. 2004. A large focus of naturally acquired Plasmodium knowlesi infections in human beings. *Lancet,* 363, 1017-24.

SINNIS, P. & SIM, B. K. 1997. Cell invasion by the vertebrate stages of Plasmodium. *Trends Microbiol,* 5, 52-8.

SMITH, J. D., CHITNIS, C. E., CRAIG, A. G., ROBERTS, D. J., HUDSON-TAYLOR, D. E., PETERSON, D. S., PINCHES, R., NEWBOLD, C. I. & MILLER, L. H. 1995. Switches in expression of Plasmodium falciparum var genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. *Cell,* 82, 101-10.

SMITH, J. D., KYES, S., CRAIG, A. G., FAGAN, T., HUDSON-TAYLOR, D., MILLER, L. H., BARUCH, D. I. & NEWBOLD, C. I. 1998. Analysis of adhesive domains from the A4VAR Plasmodium falciparum erythrocyte membrane protein-1 identifies a CD36 binding domain. *Mol Biochem Parasitol,* 97, 133-48.

SMITH, T. G., WALLIKER, D. & RANFORD-CARTWRIGHT, L. C. 2002. Sexual differentiation and sex determination in the Apicomplexa. *Trends Parasitol,* 18, 315-23.

SNOW, R. W., OMUMBO, J. A., LOWE, B., MOLYNEUX, C. S., OBIERO, J. O., PALMER, A., WEBER, M. W., PINDER, M., NAHLEN, B., OBONYO, C., NEWBOLD, C., GUPTA, S. & MARSH, K. 1997. Relation between severe malaria morbidity in children and level of Plasmodium falciparum transmission in Africa. *Lancet,* 349, 1650-4.

SPENCE, P. J., JARRA, W., LEVY, P., REID, A. J., CHAPPELL, L., BRUGAT, T., SANDERS, M., BERRIMAN, M. & LANGHORNE, J. 2013. Vector transmission regulates immune control of Plasmodium virulence. *Nature,* 498, 228-31.

SPIELMANN, T., GARDINER, D. L., BECK, H. P., TRENHOLME, K. R. & KEMP, D. J. 2006. Organization of ETRAMPs and EXP-1 at the parasite-host cell interface of malaria parasites. *Mol Microbiol,* 59, 779-94.

SPIELMANN, T. & GILBERGER, T. W. 2010. Protein export in malaria parasites: do multiple export motifs add up to multiple export pathways? *Trends Parasitol,* 26, 6-10.

SRINIVASAN, P., BEATTY, W. L., DIOUF, A., HERRERA, R., AMBROGGIO, X., MOCH, J. K., TYLER, J. S., NARUM, D. L., PIERCE, S. K., BOOTHROYD, J. C., HAYNES, J. D. & MILLER, L. H. 2011. Binding of Plasmodium merozoite proteins RON2 and AMA1 triggers commitment to invasion. *Proc Natl Acad Sci USA,* 108, 13275-80.

STAALSOE, T., SHULMAN, C. E., BULMER, J. N., KAWUONDO, K., MARSH, K. & HVIID, L. 2004. Variant surface antigen-specific IgG and protection against clinical consequences of pregnancy-associated Plasmodium falciparum malaria. *Lancet,* 363, 283-9.

STANKE, M. & MORGENSTERN, B. 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Research 22 W465-W467.*

STURM, A., AMINO, R., VAN DE SAND, C., REGEN, T., RETZLAFF, S., RENNENBERG, A., KRUEGER, A., POLLOK, J. M., MENARD, R. & HEUSSLER, V. T. 2006. Manipulation of host hepatocytes by the malaria parasite for delivery into liver sinusoids. *Science,* 313, 1287-90.

SU, X. Z., HEATWOLE, V. M., WERTHEIMER, S. P., GUINET, F., HERRFELDT, J. A., PETERSON, D. S., RAVETCH, J. A. & WELLEMS, T. E. 1995. The large diverse gene family var encodes proteins involved in cytoadherence and antigenic variation of Plasmodium falciparum-infected erythrocytes. *Cell,* 82, 89-100.

SUTHERLAND, C. J. 2001. Stevor transcripts from Plasmodium falciparum gametocytes encode truncated polypeptides. *Mol Biochem Parasitol,* 113, 331-5.

SWAIN, M. T., TSAI, I. J., ASSEFA, S. A., NEWBOLD, C., BERRIMAN, M. & OTTO, T. D. 2012. A post-assembly genome-improvement toolkit (PAGIT) to obtain annotated genomes. *Nat Protoc,* 7, 1260-84.

TAKEO, S., ARUMUGAM, T. U., TORII, M. & TSUBOI, T. 2009. Wheat germ cell-free technology for accelerating the malaria vaccine research. *Expert Opin Drug Discov,* 4, 1191-9.

TANG, J., HU, M., LEE, S. & ROBLIN, R. 2000. A polymerase chain reaction based method for detecting Mycoplasma/Acholeplasma contaminants in cell culture. *J Microbiol Methods,* 39, 121-6.

TAYLOR, H. M., KYES, S. A., HARRIS, D., KRIEK, N. & NEWBOLD, C. I. 2000. A study of var gene transcription in vitro using universal var gene primers. *Mol Biochem Parasitol,* 105, 13-23.

TEMPLETON, T. J. 2009a. The varieties of gene amplification, diversification and hypervariability in the human malaria parasite, Plasmodium falciparum. *Mol Biochem Parasitol.*

TEMPLETON, T. J. 2009b. The varieties of gene amplification, diversification and hypervariability in the human malaria parasite, Plasmodium falciparum. *Mol Biochem Parasitol,* 166, 109-16.

THAM, W. H., PAYNE, P. D., BROWN, G. V. & ROGERSON, S. J. 2007. Identification of basic transcriptional elements required for rif gene transcription. *Int J Parasitol,* 37, 605-15.

THATHY, V., FUJIOKA, H., GANTT, S., NUSSENZWEIG, R., NUSSENZWEIG, V. & MENARD, R. 2002. Levels of circumsporozoite protein in the Plasmodium oocyst determine sporozoite morphology. *EMBO J,* 21, 1586-96.

TIBURCIO, M., NIANG, M., DEPLAINE, G., PERROT, S., BISCHOFF, E., NDOUR, P. A., SILVESTRINI, F., KHATTAB, A., MILON, G., DAVID, P. H., HARDEMAN, M., VERNICK, K. D., SAUERWEIN, R. W., PREISER, P. R., MERCEREAU-PUIJALON, O., BUFFET, P., ALANO, P. & LAVAZEC, C. 2012. A switch in infected erythrocyte deformability at the maturation and blood circulation of Plasmodium falciparum transmission stages. *Blood,* 119, e172-80.

TILLEY, L., SOUGRAT, R., LITHGOW, T. & HANSSEN, E. 2008. The twists and turns of Maurer's cleft trafficking in P. falciparum-infected erythrocytes. *Traffic,* 9, 187-97.

TSAI, I. J., OTTO, T. D. & BERRIMAN, M. 2010. Improving draft assemblies by iterative mapping and assembly of short reads to eliminate gaps. *Genome Biology 11:R41.*

TSUBOI, T., TAKEO, S., IRIKO, H., JIN, L., TSUCHIMOCHI, M., MATSUDA, S., HAN, E. T., OTSUKI, H., KANEKO, O., SATTABONGKOT, J., UDOMSANGPETCH, R., SAWASAKI, T., TORII, M. & ENDO, Y. 2008. Wheat germ cell-free system-based production of malaria proteins for discovery of novel vaccine candidates. *Infect Immun,* 76, 1702-8.

TURNER, L., LAVSTSEN, T., BERGER, S. S., WANG, C. W., PETERSEN, J. E., AVRIL, M., BRAZIER, A. J., FREETH, J., JESPERSEN, J. S., NIELSEN, M. A., MAGISTRADO, P., LUSINGU, J., SMITH, J. D., HIGGINS, M. K. & THEANDER, T. G. 2013. Severe malaria is associated with parasite binding to endothelial protein C receptor. *Nature,* 498, 502-5.

URBAN, B. C., FERGUSON, D. J., PAIN, A., WILLCOX, N., PLEBANSKI, M., AUSTYN, J. M. & ROBERTS, D. J. 1999. Plasmodium falciparum-infected erythrocytes modulate the maturation of dendritic cells. *Nature,* 400, 73-7.

VANDERBERG, J. P., CHEW, S. & STEWART, M. J. 1990. Plasmodium sporozoite interactions with macrophages in vitro: a videomicroscopic analysis. *J Protozool,* 37, 528-36.

VOSS, T. S., HEALER, J., MARTY, A. J., DUFFY, M. F., THOMPSON, J. K., BEESON, J. G., REEDER, J. C., CRABB, B. S. & COWMAN, A. F. 2006. A var gene promoter controls allelic exclusion of virulence genes in Plasmodium falciparum malaria. *Nature,* 439, 1004-8.

VOSS, T. S., TONKIN, C. J., MARTY, A. J., THOMPSON, J. K., HEALER, J., CRABB, B. S. & COWMAN, A. F. 2007. Alterations in local chromatin

environment are involved in silencing and activation of subtelomeric var genes in Plasmodium falciparum. *Mol Microbiol,* 66, 139-50.

WANG, C. W., MAGISTRADO, P. A., NIELSEN, M. A., THEANDER, T. G. & LAVSTSEN, T. 2009. Preferential transcription of conserved rif genes in two phenotypically distinct Plasmodium falciparum parasite lines. *Int J Parasitol,* 39, 655-64.

WANG, C. W., MWAKALINGA, S. B., SUTHERLAND, C. J., SCHWANK, S., SHARP, S., HERMSEN, C. C., SAUERWEIN, R. W., THEANDER, T. G. & LAVSTSEN, T. 2010. Identification of a major rif transcript common to gametocytes and sporozoites of Plasmodium falciparum. *Malar J,* 9, 147.

WARIMWE, G. M., KEANE, T. M., FEGAN, G., MUSYOKI, J. N., NEWTON, C. R., PAIN, A., BERRIMAN, M., MARSH, K. & BULL, P. C. 2009. Plasmodium falciparum var gene expression is modified by host immunity. *Proc Natl Acad Sci USA,* 106, 21801-6.

WEBER, J. L. 1988. Interspersed repetitive DNA from Plasmodium falciparum. *Mol Biochem Parasitol,* 29, 117-24.

WICKHAM, M. E., RUG, M., RALPH, S. A., KLONIS, N., MCFADDEN, G. I., TILLEY, L. & COWMAN, A. F. 2001. Trafficking and assembly of the cytoadherence complex in Plasmodium falciparum-infected human erythrocytes. *Embo J,* 20, 5636-49.

WINTER, G., KAWAI, S., HAEGGSTROM, M., KANEKO, O., VON EULER, A., KAWAZU, S., PALM, D., FERNANDEZ, V. & WAHLGREN, M. 2005. SURFIN is a polymorphic antigen expressed on Plasmodium falciparum merozoites and infected erythrocytes. *J Exp Med,* 201, 1853-63.

WITMER, K., SCHMID, C. D., BRANCUCCI, N. M., LUAH, Y. H., PREISER, P. R., BOZDECH, Z. & VOSS, T. S. 2012. Analysis of subtelomeric virulence gene families in Plasmodium falciparum by comparative transcriptional profiling. *Mol Microbiol,* 84, 243-59.

ZERBINO, D. R. & BIRNEY, E. 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.,* 18, 821-829.

ZHANG, Q., SIEGEL, T. N., MARTINS, R. M., WANG, F., CAO, J., GAO, Q., CHENG, X., JIANG, L., HON, C. C., SCHEIDIG-BENATAR, C., SAKAMOTO, H., TURNER, L., JENSEN, A. T., CLAES, A., GUIZETTI, J., MALMQUIST, N. A. & SCHERF, A. 2014. Exonuclease-mediated degradation of nascent RNA silences genes linked to severe malaria. *Nature,* 513, 431-5.